Bachelor thesis

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

Image Processing Methods for Determining Honeycomb State

Dominik Eichenberger

Supervisors: Mgr. Ondřej Drbohlav, Ph.D., Ing. Jiří Ulrich
Study program: Open informatics
Specialisation: Artificial Intelligence and Computer Science
May 2026

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Eichenberger** Jméno: **Dominik** Osobní číslo: **509359**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra kybernetiky**

Studijní program: **Otevřená informatika**

Specializace: **Základy umělé inteligence a počítačových věd**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Metody zpracování obrazu pro určení stavu včelí plástve

Název bakalářské práce anglicky:

Image Processing Methods for Determining Honeycomb State

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Mgr. Ondřej Drbohlav, Ph.D. katedra kybernetiky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Ing. Jiří Ulrich centrum umělé inteligence FEL

Datum zadání bakalářské práce: **02.02.2026** Termín odevzdání bakalářské práce: ____________

Platnost zadání bakalářské práce: **19.09.2027**

podpis vedoucí(ho) ústavu/katedry

podpis proděkana(ky) z pověření děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

Datum převzetí zadání

Eichenberger Dominik

Podpis studenta

© ČVUT v Praze, Design: ČVUT v Praze, VIC

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Eichenberger** Jméno: **Dominik** Osobní číslo: **509359**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra kybernetiky**

Studijní program: **Otevřená informatika**

Specializace: **Základy umělé inteligence a počítačových věd**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Metody zpracování obrazu pro určení stavu včelí plástve

Název bakalářské práce anglicky:

Image Processing Methods for Determining Honeycomb State

Pokyny pro vypracování:

1. Familiarize yourself with standard methods of registering and stitching images in presence of moving objects that may obscure the scene background.
2. Make a literature overview of papers dealing with determining the honeycomb state from images.
3. Propose methods for determining the honeycomb state from large image data. Suggest criteria for their performance evaluation. Use the datasets available at CTU (e.g., within RoboRoyale project, https://roboroyale.eu/).
4. Implement the proposed methods and criteria.

Seznam doporučené literatury:

[1] Ulrich, J.; Stefanec, M.; Rekabi-Bana, F.; Fedotoff, L. A.; Rouček, T.; Gündeğer, B. Y.; Saadat, M.; Blaha, J.; Janota, J.; Hofstadler, D. N.; Žampachů, K.; Keyvan, E. E.; Erdem, B.; Şahin, E.; Alemdar, H.; Turgut, A. E.; Arvin, F.; Schmickl, T.; Krajník, T.: Autonomous tracking of honey bee behaviors over long-term periods with cooperating robots. Science Robotics 9(89), 2024.
[2] Torralba, A.; Isola, P.; Freeman, W.: Foundations of Computer Vision. MIT Press, 2024.
[3] Smith, R.; Self, M.; Cheeseman, P: Estimating uncertain spatial relationships in robotics.
In UAI'86: Proceedings of the Second Conference on Uncertainty in Artificial Intelligence (pp. 267–288). AUAI Press, 1986.
[4] Hartley, R.; Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, 2000.

© ČVUT v Praze, Design: ČVUT v Praze, VIC

FAKULTA ELEKTROTECHNICKÁ
FACULTY OF ELECTRICAL ENGINEERING
Technická 2
166 27 Praha 6

DECLARATION

I, the undersigned

Student's surname, given name(s): Eichenberger Dominik
Personal number: 509359
Programme name: Open Informatics

hereby declare that the Bachelor’s Thesis titled

Image Processing Methods for Determining Honeycomb State

is my own independent work and that I have declared all sources of information used in accordance with the Methodological Guideline for adhering to ethical principles when elaborating an academic final thesis and the Framework Rules for the use of artificial intelligence at CTU for study and pedagogical purposes in Bachelor and continuing Masters studies.

I declare that I have used artificial intelligence tools during the preparation and writing of the final thesis.
I've verified the generated content. I confirm that I am aware that I am fully responsible for the content of the final thesis.

In Prague on 30.04.2026

Dominik Eichenberger
....
student's signature

Acknowledgements

I would like to express my sincere gratitude to my supervisors: Mgr. Ondřej Drbohlav, Ph.D., for his expert guidance, valuable feedback, and the time he dedicated to me throughout the work on this thesis, and Ing. Jiří Ulrich, for introducing me to this fascinating topic, providing access to the RoboRoyale dataset, and for his continuous technical support. Furthermore, I am grateful to Ing. Zdeněk Rozsypálek for providing the bee detection model, which proved essential for this work. I would also like to thank my family for their continuous support and patience. Last, but not least, my special thanks go to my good friends Jiří Zeman, Eliška Folková, Adéla Stuchlíková and Kateřina Finková for keeping me sane throughout my studies and the writing process.

Abstract

This bachelor thesis addresses the challenge of autonomous visual mapping of a honeybee comb within observation hives. Monitoring the comb's state is essential for assessing colony health, yet the process is hindered by the highly dynamic environment, the repetitive structural patterns of the cells, and significant illumination variations. The proposed system employs a direct image alignment method, specifically a weighted Kanade-Lucas-Tomasi (KLT) framework applied to dense hexagonal grids. This approach enables robust transformation estimation in environments lacking distinct point features. To mitigate the influence of moving bees, which act as dynamic outliers, an original weighting scheme based on bee detection and distance transforms is introduced. Global geometric consistency of the map is maintained through pose graph optimization using Iteratively Reweighted Least Squares with a Huber loss function. Experimental results demonstrate that combining odometry-based initialization with weighted image alignment allows for the construction of accurate comb map even under challenging conditions. This work establishes the foundations for future semantic mapping and long-term automated monitoring of honeybee colonies.

Keywords: Computer vision, Image stitching, Direct image alignment, Pose graph optimization, Honeybee monitoring

Supervisors: Mgr. Ondřej Drbohlav, Ph.D., Ing. Jiří Ulrich

Abstrakt

Tato bakalářská práce se zabývá problematikou automatického vizuálního mapování plástve ve sledovaných včelích úlech. Monitoring stavu plástve je klíčový pro hodnocení zdraví včelstev. Proces je avšak komplikován dynamickým prostředím, repetitivní strukturou buněk a výraznými změnami osvětlení. Navržený systém využívá metodu přímého zarovnávání snímků (Weighted Kanade-Lucas-Tomasi) aplikovanou na hexagonální mřížku bodů, což umožňuje robustní odhad transformací i v prostředí s nedostatkem unikátních rysů. Pro potlačení vlivu pohybujících se včel, které v obraze vystupují jako outliery, je představen originální algoritmus generování vah založený na detekci včel a distance transformaci. Globální geometrická konzistence mapy je následně zajištěna optimalizací pose grafu pomocí iterativně vážených nejmenších čtverců s Huberovou ztrátovou funkcí. Experimentální výsledky ukazují, že kombinace počátečního odhadu z odometrie a váženého zarovnávání snímků umožňuje vytvoření dostatečně přesné mapy plástve i za nepříznivých podmínek. Tato práce pokládá základy pro budoucí sémantické mapování a dlouhodobý automatizovaný monitoring včelstev.

Klíčová slova: Počítačové vidění, Sešívání snímků, Zarovnávání snímků, Optimalizace pose grafu, Sledování včel

Contents

Figures

Tables

Chapter 1

Introduction

Honeybees (*Apis mellifera*) play an indispensable role in the global ecosystem, primarily as key pollinators for both agricultural crops and wild vegetation. In recent decades, however, bee colonies have faced increasing threats, such as loss of biodiversity, excessive use of pesticides, and the spread of parasites, leading to phenomena like Colony Collapse Disorder (CCD). To maintain the health of these colonies, detailed and regular monitoring of the honeycomb state, specifically the amount of brood, honey, and pollen stores, is essential. Traditional inspection methods, however, require physical opening of the hive, which stresses the bees, disrupts critical thermoregulation, and temporarily alters their natural behavior, thereby biasing observation results.

This challenge is addressed by the international RoboRoyale [1] project, within which a robotic observation platform AROBA [2] has been developed for continuous and non-invasive monitoring of honeybee colonies in observation hives. The robotic system utilizes a camera operating in the infrared spectrum to track the queen bee and her surroundings without disrupting the colony's natural light regime.

To overcome these physical constraints, the robotic platform moves a camera orthogonally across the comb. However, because this mobile setup's field of view covers only a small fraction of the surface at any given moment, constructing a unified, spatially consistent global map from these individual overlapping images is an essential prerequisite for any comprehensive biological analysis. Although the platform allows for the capture of high-resolution images, the camera's field of view covers only a small fraction of the comb surface at any given moment. While a stationary high-resolution camera could theoretically monitor the entire hive simultaneously, it introduces severe practical limitations. First, capturing the whole comb at a magnification sufficient to resolve millimeter-sized features generates an enormous volume of data, making real-time processing and storage highly impractical. Second, a static wide-angle view suffers from significant perspective distortion. The further the honeycomb cells are from the camera's optical axis, the more their interiors are occluded by the cell walls, making reliable inspection of brood and stores impossible. For comprehensive biological analysis, such as mapping queen egg-laying dynamics or the distribution of stores, it is necessary to have a unified view of the entire comb surface in the form of a spatially consistent map.

The objective of this bachelor thesis is to design and implement a robust method for the automatic construction of such a map from individual images acquired by the robotic system. This task, which may appear as a standard image stitching problem, is significantly complicated in the hive environment by the highly repetitive cell structure, varying illumination, and the constant movement of bees, which act as dynamic outliers in the images. This work presents a solution based on direct image alignment (Weighted KLT) integrated into global pose-graph optimization. The proposed approach overcomes the limitations of mechanical odometry and establishes an accurate geometric foundation for future long-term monitoring and semantic analysis of the colony's state.

Chapter 2

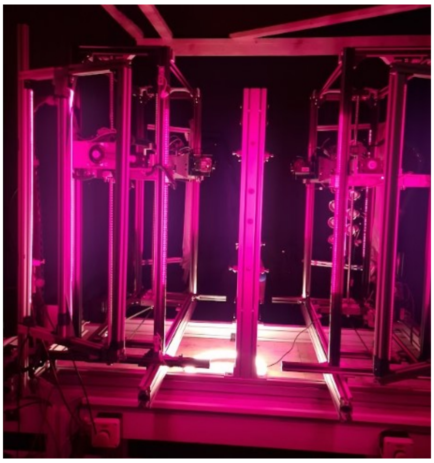
Problem overview

This chapter defines the core objective of this work: constructing a globally consistent spatial map of the honeycomb. Furthermore, it outlines the primary environmental and technical challenges encountered during the visual mapping and establishes the fundamental geometric assumptions necessary for the proposed registration pipeline.

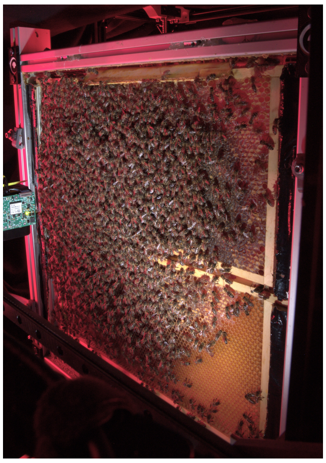
2.1 Acquisition setup

The image data utilized in this work were acquired using the robotic observation platform AROBA [2], shown in Figure 2.1. The system is designed for continuous, non-invasive monitoring of a honeybee colony and consists of several key components:

- **Observation Hives:** The colony is housed in specialized aluminum observation hives enclosed by glass panels. The internal environment is connected to the outdoors via a plastic tube, allowing the bees to forage and exhibit natural behavior while being continuously monitored.
- **Robotic Gantry System:** A Cartesian gantry manipulator is installed around the hive structure. This high-precision robotic framework carries a movable camera unit, allowing it to navigate across the entire surface of the comb. The physical movements of the gantry provide the raw mechanical odometry used as the initial guess for our image registration pipeline.
- **Vision and Illumination System:** The imaging unit is equipped with a motorized zoom lens capable of achieving a spatial resolution ranging from 16 to $67 \mu m$ per pixel [2]. This high magnification is essential for resolving minute intra-cell structures, such as honeybee eggs (approximately 1×0.3 mm). To prevent disruption of the colony's natural behavior and circadian rhythms, all visual observations are conducted under near-infrared (NIR) illumination at a wavelength of 750 nm [3].
- **Scanning Protocol and Scheduling:** The full-comb mapping is executed using a sequential, overlapping grid scanning pattern (serpentine



(a) : AROBA observation system



(b) : Observation hive

Figure 2.1: The robotic data acquisition platform developed within the RoboRoyale project. (a) The high-precision Cartesian gantry system AROBA, (b) The glass-walled observation hive housing the monitored honeybee colony.

scan). A single complete pass over the hive surface generates approximately 8000 to 9000 individual image tiles. Since the primary objective of the robotic platform is to continuously track the queen bee, the full-comb mapping serves as a secondary task. These extensive scans are scheduled opportunistically during periods when the queen is resting, ensuring that crucial behavioral data is not lost while acquiring the static map of the comb.

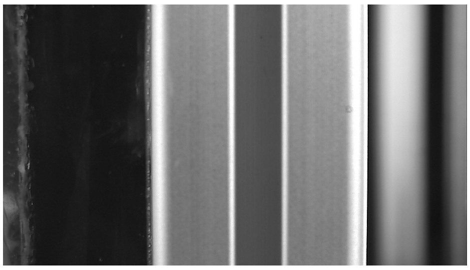
2.2 Challenges

The construction of a global honeycomb map is complicated by several interlocking factors:

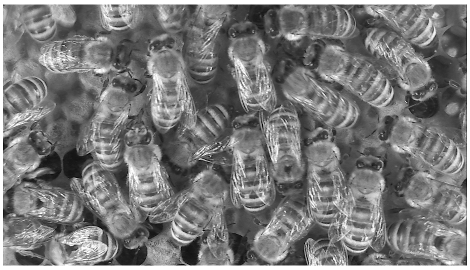
First, the camera trajectory is not always uniform, occasionally moving over regions with minimal distinctive structure. In such instances, the estimation of displacement relies heavily on the system's odometry, which is inherently subject to cumulative drift over time. Figure 2.2a illustrates a typical scenario during a vertical motion phase. The image contains only homogeneous surfaces and low-frequency gradients, making the robust detection and matching of distinctive visual keypoints impossible. In such instances, the system is fundamentally reliant on odometry for vertical motion estimation.

Second, the honeycomb consists of highly self-similar hexagonal cells. This repetitive geometry makes it difficult to establish unique correspondences and increases the risk of the alignment converging to a local minimum (i.e., "snapping" to an adjacent cell).

Furthermore, lighting conditions across the hive surface are often non-uniform. Global brightness gradients and local shadows within cells introduce intensity variations between images taken from different perspectives, as



(a) : Low-texture environment during vertical displacement.



(b) : Severe dynamic occlusion by a dense bee cluster.

Figure 2.2: Primary visual registration challenges. (a) Scenario observed during vertical displacement (y-axis) where lack of texture prevents keypoint matching. (b) Dense clusters of bees occluding the static honeycomb background.

shown in Figure 2.3. These variations pose a significant challenge for direct intensity- or gradient-based registration methods.

This issue is further complicated by the serpentine scanning pattern, which yields high spatial redundancy intra-row but low overlap inter-row. The extensive length of the scanning trajectory introduces a significant time delay before the camera revisits a specific x-coordinate in a subsequent row. This temporal gap allows for substantial bee movement and odometry drift to accumulate, as demonstrated in Figure 2.4. Consequently, computing reliable relative transformations between adjacent rows becomes the primary bottleneck in maintaining global geometric consistency.

Finally, in certain regions of the hive, such as the brood nest or the dance floor, bees congregate in dense, multi-tiered layers to regulate temperature or communicate. As illustrated in Figure 2.2b, these clusters can largely occlude the underlying honeycomb structure. From a computer vision perspective, this removes all usable static background information within the overlapping region. If a standard visual registration algorithm is applied here, it will inevitably track the movement of the bees rather than the comb, leading to a misalignment. In extreme cases of near-total occlusion, image-based registration becomes mathematically ill-posed, and the mapping system must rely entirely on the underlying mechanical odometry to bridge the gap.

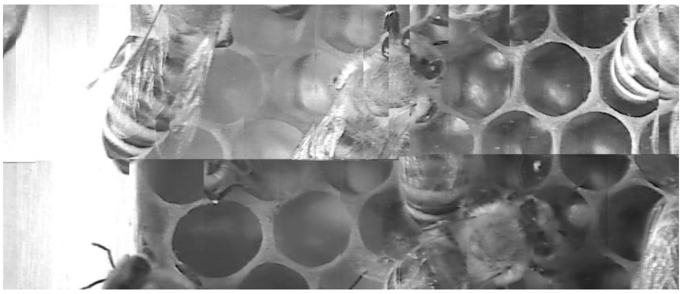


Figure 2.3: Illustration of ilumination and temporal challenges, showcasing an overlap region between two adjacent rows in the serpentine scan.

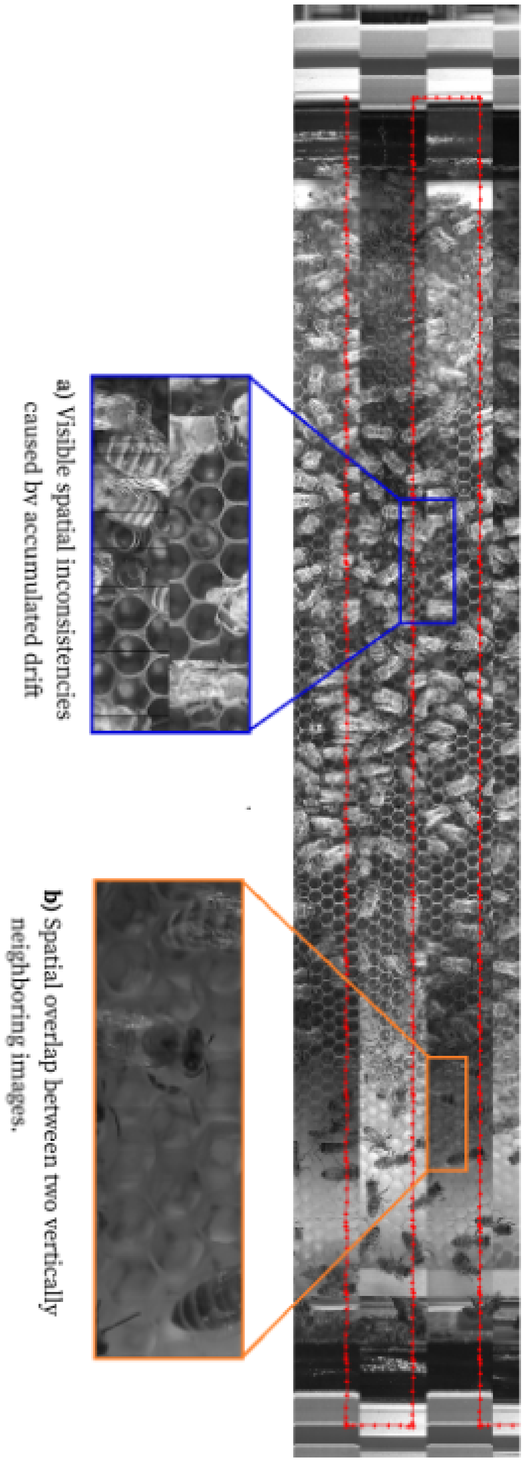


Figure 2.4: Visualization of the serpentine scanning trajectory highlighting the core registration challenges. **(Red)** Trajectory of the camera system. **(Blue)** A detailed view of the temporal challenge; the delay between acquiring overlapping regions results in drift and severe dynamic occlusions, visible as "ghosting" of the moving bees, which disrupts standard visual registration. **(Orange)** A detailed view of the spatial overlap area between two vertically neighboring images, illustrating the available structural information.

2.3 Assumptions

A critical feature of the proposed mapping pipeline is that it does not require perfect observation conditions across the entire hive surface. Because the system utilizes global pose-graph optimization with robust loss functions, individual pairwise registration failures are explicitly expected and tolerated. Instead, the approach fundamentally relies on a specific set of geometric and environmental assumptions holding true for a statistically sufficient subset of the overlapping regions, ensuring that the final optimized graph remains fully connected.

First, the camera motion is assumed to be strictly planar and parallel to the hive surface, allowing the honeycomb to be treated as a 2D plane. Although biological variations in the comb, such as protruding capped honey versus recessed brood cells, locally violate this assumption by introducing depth variations, these differences are small enough relative to the camera distance to be adequately approximated and compensated for by the 6-parameter affine transformation model.

Second, despite the hive being a highly dynamic environment, the algorithm assumes static background dominance within successfully registered overlapping regions. This implies that the structurally static elements, namely the hexagonal cell walls, provide sufficient gradient information to outweigh the moving outliers caused by bee movement. If a local area is completely occluded, the pairwise registration will fail and be discarded, assuming such extreme occlusions are spatially localized and do not completely sever the connectivity between adjacent rows.

Furthermore, the system relies on odometry to provide a locally reliable prior for the majority of the robotic movements. To achieve successful alignment, the initial displacement guess must fall within the convergence basin of the pyramidal Kanade-Lucas-Tomasi (KLT) [4] tracker, which typically corresponds to half of a cell diameter. Isolated instances where severe odometry jumps violate this assumption are mathematically treated as outliers and suppressed by the Huber loss during global optimization. Ultimately, the overarching assumption of this work is that the serpentine scanning trajectory provides sufficient spatial overlap.

2.4 State of the art

Several methodologies and robotic systems have been proposed for automated honeybee monitoring and image-based hive analysis. Recent advances in robotic beekeeping demonstrate a growing interest in long-term autonomous monitoring of honeybee colonies. For instance, Barmak et al. [5] proposed an automated robotic hive platform capable of actively regulating hive conditions to improve colony survival during winter periods. More recently, the RoboRoyale project introduced a robotic observation system for continuous autonomous queen tracking and behavioral analysis inside a live colony [2].

This system employs cooperating robotic agents operating directly within the hive environment to continuously acquire large-scale visual data with minimal disturbance to the bees.

When considering the spatial alignment of such visual data, traditional feature-based registration pipelines often rely on sparse keypoint detectors and local descriptors such as SIFT [6], SuperPoint [7], or LoFTR [8]. While highly effective in general computer vision applications, these approaches frequently struggle in honeybee comb environments due to the highly repetitive hexagonal structure and the lack of uniquely distinguishable visual features. Consequently, feature-based methods often achieve significantly lower precision than direct approaches in such scenarios [9].

To mitigate these limitations, direct correlation-based methods estimate transformations by maximizing photometric consistency directly between overlapping image regions. Techniques such as normalized cross-correlation, phase correlation, or KLT optimization [4] can provide high alignment precision even in repetitive environments, particularly when combined with a reliable initialization from odometry. However, they remain highly sensitive to changes in illumination, dynamic occlusions caused by bee movement, and accumulated drift over long trajectories. Conversely, in robotic systems with precise mechanical motion control, odometry-centered mapping can provide a strong baseline estimate for spatial reconstruction. As discussed by Janota et al. [9], sufficiently accurate odometry may even outperform purely vision-based methods in highly repetitive scenes where visual alignment risks converging to incorrect local minima. Nevertheless, odometry alone is generally insufficient for maintaining long-term global consistency due to the inevitable accumulation of systematic drift errors, which necessitates the integration of global optimization frameworks.

Chapter 3

Problem formulation

This chapter mathematically formalizes the mapping task. It defines the structure of the input data provided by the robotic platform, specifies the exact format of the expected output, and establishes the objective function used to optimize global spatial consistency

3.1 Input

The original input is provided in the form of a timestamped series of photos of the honeybee hive and a series of periodically reported timestamped coordinates of the camera's position. The camera system moves in a serpentine pattern on a plane approximately parallel to the hive and takes grayscale photos with a resolution of 1920×1080 pixels. Formally, the input can be described as $\{(img_i, t_i)\}_{i=1}^{N}$, $\{((x_j, y_j), t_j)\}_{j=1}^{M}$. , where N is the number of images taken during a scan and M represents the number of coordinates reported by the acquisition system. As we need at least approximate coordinates for each image, the input is simplified using interpolation into $\{img_i, (x_i, y_i)\}_{i=1}^{N}$

3.2 Output

The output of the proposed method is a globally consistent spatial representation of the observed hive surface constructed from the input image sequence. Formally, the output consists of a set of images and their estimated global transformations $\{(T_i, img_i)\}_{i=1}^{N}$, where each transformation T_i is a 2D affine transformation (containing 6 degrees of freedom) that maps image img_i from its local coordinate system into a common global coordinate frame. These transformations are obtained through graph-based optimization and allow for the mutual alignment of all images in the sequence. Based on the estimated transformations, it is further possible to construct a composite mosaic of the hive surface, which provides a unified view of the scanned area. The resulting representation is designed to be incrementally extendable, allowing us to update the map with additional images from future scans.

3.3 Objective Definition

Given the input image sequence and approximate camera positions, the objective is to estimate a set of global image transformations $\{T_i^*\}_{i=1}^N$ such that the resulting arrangement of all images is globally geometrically consistent.

Let $\mathcal{O} = \{(i, j)\}$ denote the set of overlapping image pairs. For each pair $(i, j) \in \mathcal{O}$, a relative transformation T_{ij} is estimated using pairwise image registration.

The global optimization problem can then be formulated as minimizing the inconsistency between the estimated global transformations and the observed pairwise transformations:

$$\{T_k^*\}_{k=1}^N = \arg min_{\{T_k\}_{k=1}^N} \sum_{(i,j)\in\mathcal{O}} d(T_{ij}, T_j T_i^{-1}) \tag{3.1}$$

where $d(\cdot, \cdot)$ denotes a distance measure between two transformations. In our context, this distance is not evaluated directly on the transformation matrices, but rather as the spatial Euclidean distance between a set of reference points mapped by these transformations. This ensures a physically meaningful minimization of pixel misalignment in the image domain. This formulation corresponds to a pose-graph optimization problem [10], where nodes represent images and their global poses, while edges represent pairwise relative transformations estimated from image alignment.

Chapter 4

Methodology

This chapter provides a description of the proposed technical framework for autonomous honeycomb mapping. It begins by discussing the initial investigation into feature-based registration and the motivations for transitioning to a dense alignment approach. The core of the methodology focuses on the implementation of the Weighted Inverse Compositional KLT tracker and the original bee suppression scheme. The chapter formalizes the construction and optimization of the global pose graph.

4.1 Initial approach

Prior to adopting the direct alignment framework, an alternative registration approach based on explicit honeycomb geometry was investigated. The method utilized the Hough transform [11] to detect hexagonal cells and extract their walls. For each detected cell, a geometric descriptor was constructed by mapping the boundary traversal distance to the wall orientation angle relative to the image x-axis. Relative image displacement was then estimated by matching these descriptors between neighboring frames.

Despite promising preliminary results under controlled conditions, the method proved insufficiently robust for real hive environments. Reliable Hough line detection required hard gradient thresholds, which were highly sensitive to non-uniform illumination. Furthermore, partial occlusions caused by bees frequently disrupted the continuity of cell walls, producing fragmented or corrupted descriptors. As a result, the matching process often failed by confusing cell structures with bee edges or wing boundaries, which would now be mitigated by employing bee-based weights.

Although ultimately discarded for pairwise registration, this investigation highlighted the fundamental limitations of sparse feature-based geometric matching in highly dynamic hive environments. The observed failure modes directly motivated the dense weighted alignment approach proposed in this work, which avoids explicit feature extraction entirely.

4.2 Overview of the pipeline

The proposed method processes a sequence of images of a honeybee hive and estimates their globally consistent spatial arrangement without relying on explicit feature tracking. First, relative transformations between pairs of overlapping images are estimated directly from the image data. These pairwise transformations capture the geometric relationship between neighboring images. Next, a graph structure is constructed, where each node represents a global transformation from the origin of an image, and edges represent pairwise transformations. Final global transformations of all images are then obtained by solving a least squares optimization problem, enforcing consistency across the entire graph. Figure 4.1 illustrates the complete processing pipeline.

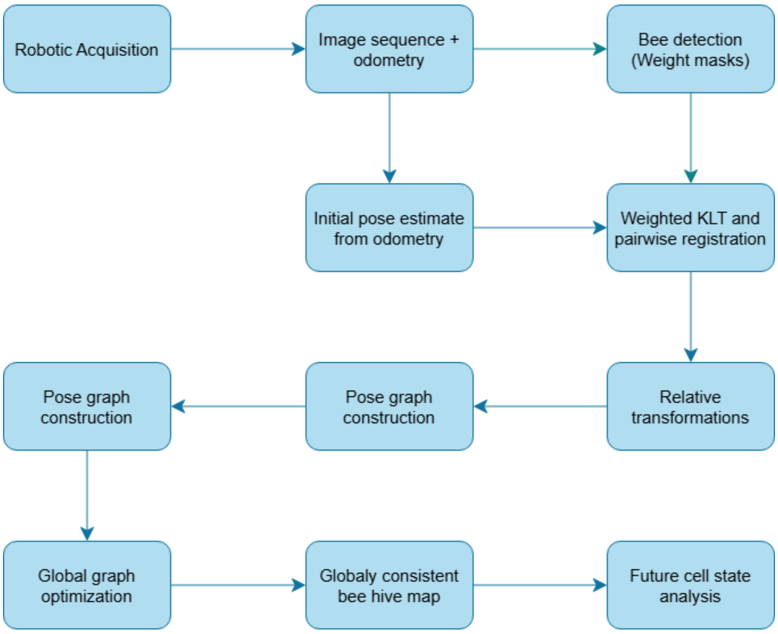


Figure 4.1: Pipeline diagram

4.3 Pairwise transformation estimation

The core of the mapping pipeline lies in the estimation of accurate relative transformations T_{ij} between overlapping image pairs. This process represents the local registration phase, where the goal is to determine the precise geometric relationship between two frames independently of the global context. Given the specific challenges of the hive environment, this task is divided into two interdependent components: a direct intensity-based alignment framework and a specialized weighting scheme designed to neutralize the influence of moving bees. The following subsections describe the mathematical formulation of these components and their integration into a robust registration unit.

4.3.1 Affine transformation

In this work, the relationship between overlapping images is modeled using a 2D affine transformation T. Unlike a simple translation, the affine model accounts for rotation, scaling, and shearing, providing the necessary flexibility to compensate for slight misalignments between the camera plane and the honeycomb surface. An affine transformation in 2D is uniquely defined by 6 parameters and can be expressed in homogeneous coordinates as:

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

where t_x, t_y represent translation, and the sub-matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ encapsulates rotation, axis scaling, and shear. [12].

4.3.2 Weighted KLT

To estimate the relative transformation T_{ij} between an image pair, we employ an intensity-based image alignment method derived from the KLT framework [4]. The classical formulation minimizes the sum of squared photometric errors between a template image I_{ref} and a target image I_{tar}:

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \sum_{\mathbf{x}} \left\| I_{tar}(W(\mathbf{x}, \mathbf{p})) - I_{ref}(\mathbf{x}) \right\|^2 , \tag{4.1}$$

where $I(\mathbf{x})$ corresponds to the monochromatic intensity of the image at coordinates $\mathbf{x}$ and $W(\mathbf{x}, \mathbf{p})$ denotes a parametric transformation of coordinates $\mathbf{x}$. Standard KLT implementations typically track small, feature-rich neighborhoods. However, due to the lack of distinct point features in the observed environment, we perform alignment over the entire image area by matching dense hexagonal grids.

Since the objective function is highly non-convex, standard gradient-based optimization converges only for small initial displacements. To handle the varying magnitudes of inter-frame motions inherent to our serpentine scanning pattern, we utilize a coarse-to-fine pyramidal approach with a scale factor of 0.5. Specifically, we employ a 3-level pyramid to accommodate the larger expected offsets during vertical transitions between adjacent rows, whereas a single level is sufficient for estimating the highly overlapping horizontal edges. Furthermore, the optimization is initialized with a starting translation $\mathbf{p}_{init}$ derived from the camera's odometry to ensure robust convergence. Figure 4.2 displays the process of KLT optimization on two neighboring images. To account for the non-static nature of the scene caused by bee movement, we extend the objective function with a per-pixel weighting term $w_k \in [0, 1]$

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} \sum_{k=1}^{N} w_k \left\| I_{tar}(W(\mathbf{x}_k, \mathbf{p})) - I_{ref}(\mathbf{x}_k) \right\|^2 , \tag{4.2}$$

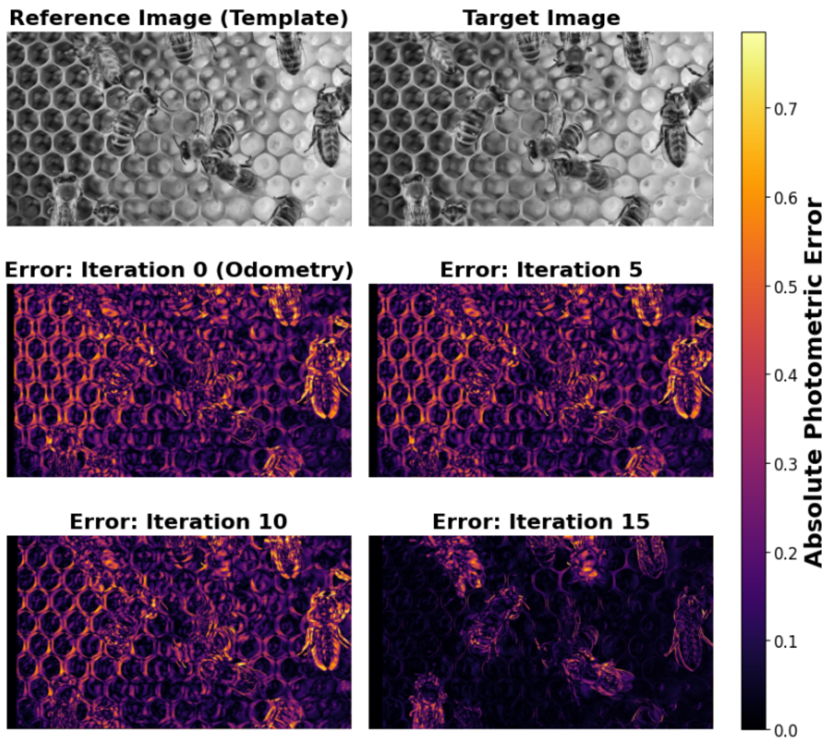


Figure 4.2: Evolution of the absolute photometric error during KLT optimization. At iteration 0 (initial odometry guess), the structural misalignment causes high error magnitudes (bright regions) along the cell edges. As the optimization progresses, the error map converges toward zero (dark regions), indicating accurate geometric alignment. The remaining areas of high error in the final iteration exclusively correspond to moving bees, which are effectively neutralized by the weighting.

The weights are designed to suppress the influence of bee movement in error calculation, specifically, lower values are assigned to regions containing detected bees. This modification ensures that the resulting transformation parameters $\mathbf{p}^*$, which define the relative transformation T_{ij}, are computed primarily from the static background. The generation of these weight maps is detailed in the next section.

4.3.3 Bee suppression through weight maps

To mitigate the impact of bee movement on the transformation estimation, we generate per-pixel weight maps $w \in [0, 1]$. These maps assign lower importance to regions occupied by bees. The generation process for an image I consists of the following steps:

1. **Detection:** The initial identification of bees is performed using a convolutional neural network. The core architecture is based on the markerless tracking model proposed by Bozek et al. [13]. However, since the original model was designed for stationary observation hives, it was subsequently

fine-tuned on a custom dataset of near-infrared (NIR) images captured directly by our robotic platform. This adaptation ensures robust performance under our specific acquisition conditions. The customized detector generates a confidence heatmap H_{bee}, where higher values indicate a high probability of a bee's presence.

2. **Masking:** The heatmap is thresholded to create a binary mask M, where $M(\mathbf{x}) = 1$ if $H_{bee}(\mathbf{x}) > 0.2$, and 0 otherwise.

3. **Distance Transform:** To create a smooth transition around detected objects, we compute the Euclidean distance transform on the inverted mask. The distances are clipped at a maximum radius of $d_{max} = 50$ pixels and normalized to the range $[0, 1]$.

4. **Weight Calculation:** The final weights are derived by squaring the normalized distances, which further suppresses contributions from pixels in close proximity to bees while maintaining a smooth gradient for the optimizer.

When estimating the transformation between a pair of images (I_i, I_j), the weight maps w_i and w_j must be combined. Since the target image I_j is warped during the optimization, its corresponding weight map w_j is likewise warped using the initial transformation parameters $\mathbf{p_{init}}$. The final joint weight for each pixel is then determined as the minimum of the two:

$$w(\mathbf{x}) = \min\left(w_i(\mathbf{x}), w_j(W(\mathbf{x}, \mathbf{p}))\right) \tag{4.3}$$

Using the minimum operator ensures that a pixel is suppressed if it is identified as part of a bee in either the reference or the warped target image. The process of weight creation is shown in Figure 4.3

4.4 Graph construction

In our graph, the nodes represent individual images paired with their global affine transformations from the system origin. Edges denote the affine camera transformations between two images. Initially, all node transformations are defined as pure translations derived from the camera system's odometry. As the camera moves in a serpentine pattern, nodes with similar y-coordinates are grouped into distinct rows. For each node, we then compute relative transformations for N_r adjacent neighbors within the same row (both on the left and right). Additionally, we find N_i nodes from the adjacent upper and lower rows that are closest in terms of their x-coordinates and try to establish edges between them. The resulting graph takes the form of set $\mathcal{O}$ denoting all pairs of images where relative transformation was successfully found, along with the set of known transformations $\{T_{ij}\}_{(i,j)\in\mathcal{O}}$ and initial global transformations of all images $\{T_i\}_{i=0}^{N}$.

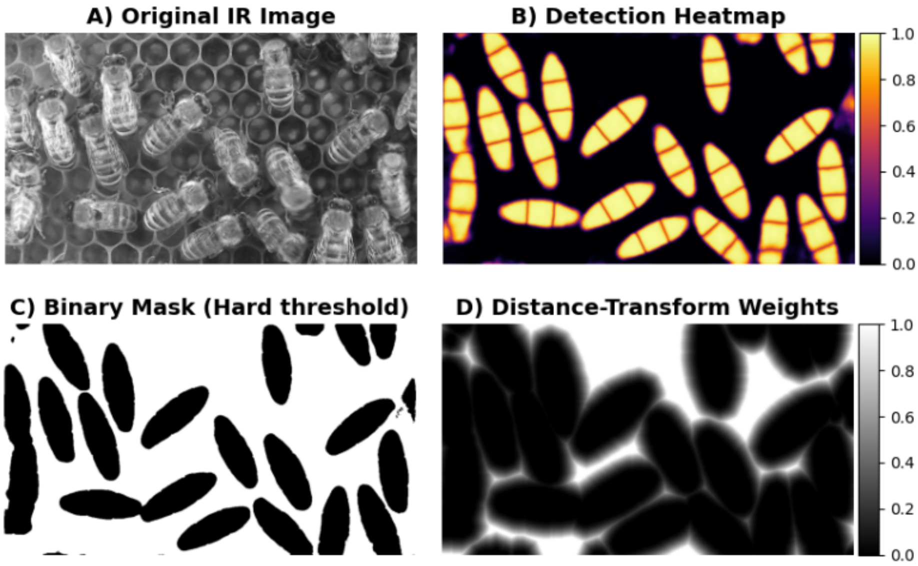


Figure 4.3: The proposed bee suppression pipeline. **(A)** The original infrared input image. **(B)** The confidence heatmap generated by the detection model. The adjacent colorbar indicates the network's prediction confidence, ranging from 0 (no bee detected) to 1 (high certainty). **(C)** A hard binary mask thresholded at 0.2 (representing standard outlier removal). **(D)** The final distance-transform weight map. The grayscale colorbar represents the continuous spatial weight assigned to each pixel during registration, scaling from 0 (completely suppressed) to 1 (fully utilized). Notice how the continuous gradient in (D) provides a much smoother optimization landscape for the KLT solver compared to the sharp discontinuities of the binary mask in (C).

4.5 Graph optimization (Least squares)

The global registration problem is formulated as a cost minimization over the set of global transformations $\{T_i\}_{i=1}^N$. A naive approach minimizes the difference between global transformations and measured relative constraints T_{ij}

$$\{T_k^*\} = \arg\min_{\{T_k\}} \sum_{(i,j)\in\mathcal{O}} \|T_j - T_{ij}\, T_i\|^2 \,. \tag{4.4}$$

However, such a direct comparison of transformation parameters often lacks a clear geometric interpretation, as it treats rotation and translation components identically. To address this, we evaluate the error based on the displacement of a set of reference points $\mathbf{p}_k$ in the image domain:

$$\{T_k^*\}_{k=1}^N = \arg\min_{\{T_i\}} \sum_{(i,j)\in\mathcal{O}} \rho\left(\sum_{i=1}^{K} \|T_j p_k - T_{ij}\, T_i p_k\|^2\right), \quad \text{s.t} \quad T_1^* = T_1. \tag{4.5}$$

where $\{\mathbf{p}_k\}_{k=1}^K$ represents a fixed grid of $K = 9$ points. This formulation ensures that the optimization minimizes the actual pixel-wise misalignment between overlapping images. The constraint $T_1^* = T_1$ fixes the global reference frame of the optimization problem. Without this constraint, the solution would remain ambiguous up to an arbitrary global transformation (i.e rotation), since applying the same transformation to all poses does not change the relative alignment error. The choice of $K = 9$ points, arranged in a 3×3 grid, represents a balanced compromise between geometric fidelity and computational efficiency. While a denser grid would provide a more granular sampling of the transformation's effect, nine points are sufficient to uniquely and robustly constrain all six degrees of freedom of the affine model. The function $\rho(\cdot)$ denotes the Huber loss [14], which enhances the robustness of the optimization against outliers typically caused by erroneous pairwise estimations. While the problem without a robust loss could be solved via linear least squares, the inclusion of $\rho(\cdot)$ necessitates an iterative approach. We employ Iteratively Reweighted Least Squares (IRLS) [15] to obtain the optimal global poses $\{T_i^*\}$.

4.6 Implementation details

While the previous sections established the theoretical and mathematical foundations of the mapping system, the practical realization of these algorithms requires consideration of numerical stability and computational performance. The implementation is designed to handle large-scale datasets efficiently within a Python-based environment. This section outlines the key technical decisions made during development.

4.6.1 Weighted Inverse Compositional KLT Implementation

To achieve both high performance and sub-pixel accuracy, we implemented a custom coarse-to-fine pyramidal KLT tracker. Rather than using the standard forward-additive formulation, our implementation utilizes the **Inverse Compositional (IC)** algorithm proposed by Baker and Matthews [4].

Computational Efficiency: In the IC formulation, the template image I_{ref} remains static while the parametric warp $W(\mathbf{x}, \mathbf{p})$ is applied to the target image I_{tar}. This allows us to evaluate the image gradients ∇I_{ref}, the Steepest Descent images, and the inverse of the Hessian matrix $\mathbf{H}^{-1}$ exactly once per pyramid level (prior to the optimization loop). Consequently, the iterative updates $\Delta\mathbf{p}$ are computed using simple matrix multiplications, significantly accelerating the pipeline compared to standard implementations.

Integration of Dynamic Weights: To mitigate the influence of bees, the precomputed weight maps are seamlessly integrated into the IC optimization. For a given image pair, we construct a joint valid mask by taking the pixel-wise minimum of the reference weight map and the warped target weight map. To correctly solve the weighted least-squares problem, the Steepest Descent images and the photometric error vectors are element-wise multiplied by the square root of the joint weights ($\sqrt{w_k}$). This efficiently zeroes out gradients originating from occlusions before they enter the Hessian formulation.

Pyramidal Execution and Interpolation:

1. **Preprocessing:** To combat the varying illumination conditions within the hive, all input images are preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE) [16] before constructing the pyramid layers.
2. **Sub-pixel Warping:** During the iterative loop, the target image must be dynamically warped. To maintain differentiability and sub-pixel accuracy, we employ bilinear interpolation via the `scipy.ndimage.map_coordinates` backend.
3. **Coarse-to-fine Tracking:** The optimization is initialized with the scaled odometry guess at the coarsest level. The resulting affine parameters are recursively scaled up and passed as the initial guess to the subsequent finer levels.

Robust Convergence Validation

Relying solely on the magnitude of the parameter update ($\|\Delta\mathbf{p}\|$) is insufficient to guarantee correct alignment in highly repetitive environments. Therefore, a post-optimization validation step is employed. The final warped target image is compared against the reference image using a weighted gradient-magnitude error metric (computed via Sobel operators). Alignments exceeding an empirically defined threshold are rejected, preventing severe misalignments from entering the global pose graph. Formally, a pairwise transformation estimation is considered successful if and only if it satisfies the following convergence criteria:

$$(iter < MAX\_ITER) \wedge (E_{grad} < \epsilon) \tag{4.6}$$

where $iter$ is the number of iterations required for convergence, $MAX\_ITER = 50$ is the iteration limit, E_{grad} is the final weighted gradient-magnitude error, and $\epsilon = 0.25$ is the acceptance threshold. If a registration fails to meet these conditions, it is treated as an outlier and is excluded from the subsequent pose-graph construction to maintain global map integrity. The weighted gradient-magnitude error E_{grad} is defined as:

$$E_{grad}(k,l) = \frac{\sum_{i=1}^{N} w_i \left|\mathcal{G}(Img_k)_i - \mathcal{G}(W(Img_l, T_{kl}))_i\right|}{\sum_{i=1}^{N} w_i}, \tag{4.7}$$

where N is the number of pixels in the overlapping region of two images, $\mathcal{G}(\cdot)$ is the gradient magnitude operator, $W(\cdot, T_{kl})$ denotes image warping using the relative transformation T_{kl}, $(Img_k)_i$ denotes the i-th pixel in image k, and w_i are per-pixel weights that suppress regions occupied by bees, computed using the procedure described in Section 4.3.3.

4.6.2 Custom Graph Optimization and Analytical Jacobian

The core of global optimization relies on solving a non-linear least squares problem. To maximize computational efficiency, we do not rely on numerical differentiation tools. Instead, we explicitly formulate and construct the analytical Jacobian matrix for the entire pose graph.

Due to the nature of our objective function, the resulting Jacobian is highly sparse. To handle this efficiently in Python, the matrix is incrementally constructed using the `scipy.sparse.lil_matrix` (List of Lists) format, which allows for fast structure modification during assembly. Once complete, it is converted into a Compressed Sparse Row (CSR) format. The optimization loop employs a custom implementation of the Levenberg-Marquardt algorithm [17, 18, 19]. While the Jacobian formulation, error calculations, and damping updates are natively implemented in our pipeline [12], the most computationally intensive step, solving the sparse linear system $\mathbf{J^T J x} = \mathbf{J^T b}$ in each iteration, is delegated to the SciPy sparse linear algebra backend [20].

4.6.3 Robust Estimation and Affine Regularization

To handle significant outliers arising from odometry drift and dynamic scene elements, two critical regularization mechanisms were implemented within the optimization loop:

- **Dynamic Huber Loss:** We utilize a robust Huber loss function to down-weight the influence of large residual errors. Rather than using a fixed threshold δ, we calculate it dynamically in each iteration using the Median Absolute Deviation (MAD) of the residuals. This adaptive scaling ensures that the threshold remains statistically sound regardless of the current magnitude of the alignment errors.

- **Affine Regularization:** Because we employ an affine transformation model, there is a risk of physically implausible scale or shear distortions accumulating over the graph. To mitigate this, a regularization block is appended to the Jacobian. Weighted by an empirically determined hyperparameter λ_{aff}, this term severely penalizes non-rigid components (scale and shear). This effectively forces the algorithm to behave predominantly as a rigid registration (translation and rotation) while retaining the minor flexibility of the affine model necessary to compensate for slight camera perspective changes.

Chapter 5

Experiments

The following chapter presents an experimental evaluation of the proposed system using real-world data acquired from the AROBA platform. It defines the characteristics of the dataset, including the distinction between intra-row and inter-row image pairs, and establishes the photometric metrics used for quantitative assessment. The experiments are designed to analyze the influence of graph connectivity, evaluate the effectiveness of the distance-transform weighting, and determine the physical limits of the system under severe visual occlusion.

5.1 Experimental Setup

This section establishes the framework for the experimental evaluation of the proposed mapping system. It provides a detailed description of the dataset used for testing, which comprises real-world image sequences captured under challenging hive conditions. Furthermore, it defines the quantitative evaluation metrics based on photometric consistency, which serve as the primary measure for assessing the geometric accuracy and registration stability of the reconstructed maps.

5.1.1 Dataset

The image data utilized in this work were acquired at the University of Graz using the robotic observation platform AROBA [2]. The data collection took place between 15 August and 9 September 2024. The experimental evaluation was conducted on a comprehensive dataset comprising approximately 16,000 image pairs, denoted as $\mathcal{P}$. These images were extracted from multiple robotic scans covering the entire surface of the observation hive.

The dataset $\mathcal{P}$ is explicitly partitioned into disjunct inter-row pairs ($\mathcal{P}_v$, connecting adjacent rows) and intra-row pairs ($\mathcal{P}_h$, sequential horizontal neighbors). This distinction is critical due to the significant disparity in registration difficulty dictated by the serpentine scanning trajectory. While intra-row pairs ($\mathcal{P}_h$) are strongly constrained by the continuous mechanical movement and provide minimal temporal delay, the inter-row pairs ($\mathcal{P}_v$) suffer from severe illumination shifts and dynamic occlusion over time. Consequently, while both subsets are utilized to evaluate the fundamental robustness of the pairwise registration algorithm, the assessment of long-range drift and global map consistency relies primarily on the structurally demanding inter-row constraints.

5.1.2 Evaluation Metric

The quality of the resulting map is evaluated using a photometric consistency measure computed on pairs of overlapping images. Only image pairs capturing the honeycomb structure are considered; images containing the hive border or the camera frame are excluded, as their alignment relies primarily on odometry rather than visual information.

For each evaluated image pair (Img_k, Img_l), the relative transformation T_{kl} is obtained from the optimized global poses. The second image is warped into the coordinate frame of the first image, and the difference in gradient magnitudes is evaluated over the overlapping region. The evaluation metric is defined as:

$$E = \frac{1}{|\mathcal{P}_v|} \sum_{(k,l)\in\mathcal{P}_v} E_{grad}(k,l) \tag{5.1}$$

where $\mathcal{P}_v$ denotes the set of evaluated image pairs and E_{grad} is defined in Equation 4.7. Lower values of E correspond to better spatial consistency between overlapping images.

5.2 Influence of Graph Connectivity

This experiment evaluates the influence of graph connectivity on the global consistency of the reconstructed map. For each image, the graph construction process connects the node with the N_i nearest neighboring images from adjacent rows, determined according to their horizontal spatial proximity. Increasing N_i therefore increases the number of inter-row constraints and improves the redundancy of the pose graph. The experiment was performed for values of $N_i \in \{0, \ldots, 9\}$, where $N_i = 0$ introduces no inter-row edges, leaving the system to rely largely on odometry.

For each configuration, the graph was optimized using the method described in Section 4.5, and the resulting map quality was evaluated using the average photometric consistency metric defined in Equation (5.1)

Table 5.1: Influence of inter-row connectivity on the reconstruction quality.

N_i	Mean E_{grad}	Median E_{grad}	Mean improvement [%]
0	0.1689	0.1658	0.0
1	0.1563	0.1542	7.4
2	0.1557	0.1543	7.8
3	0.1554	0.1537	8.0
4	0.1550	0.1535	8.2
5	0.1549	0.1536	8.3
6	0.1549	0.1533	8.3
7	0.1547	0.1536	8.4
8	0.1545	0.1532	8.5
9	0.1545	0.1533	8.5

As Table 5.1 shows, the largest improvement is achieved when introducing only the closest inter-row neighbors. Additional constraints further improve the result only marginally, suggesting that the pose graph becomes sufficiently constrained already for low values of N_i. This outcome is highly desirable from a computational standpoint, as estimating inter-row transformations requires significantly more iterations due to challenging overlaps and bee movement.

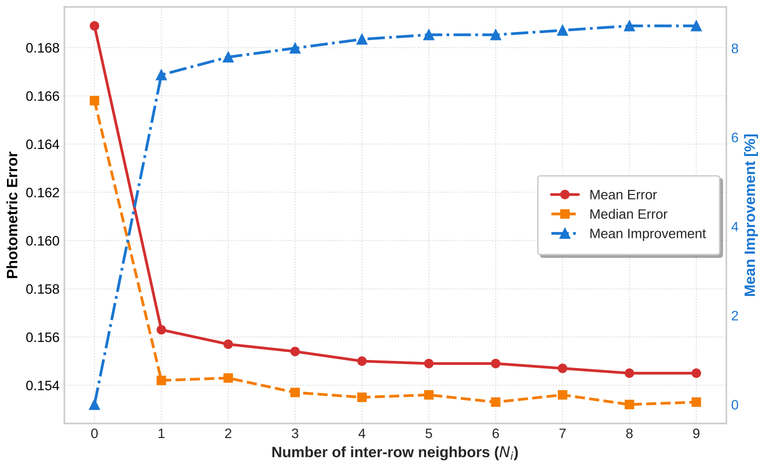


Figure 5.1: Influence of inter-row connectivity on map consistency. The photometric error drops significantly after establishing connections with the closest adjacent row ($N_i = 1$) and quickly flattens, demonstrating that redundant inter-row constraints yield diminishing returns.

5.3 Influence of Bee Weighting

This experiment evaluates the influence of different pixel weighting strategies during pairwise image registration. The following weighting schemes were compared:

1. No weighting (standard KLT),
2. Binary bee masking,
3. Distance-transform-based weighting.

The binary masking approach fully suppresses pixels classified as bees, while the distance-transform weighting gradually decreases the influence of pixels near bee regions. The latter approach provides a smoother optimization landscape and reduces sensitivity to segmentation inaccuracies near bee boundaries. In this experiment, instead of evaluating the final photometric consistency metric, we measure the KLT algorithm's success rate to estimate relative transformations across all evaluated vertical image pairs $\mathcal{P}_v$. A registration is defined as successful if the optimization converges in fewer than $MAX\_ITER = 50$ iterations and the resulting weighted gradient-magnitude error is less than $\epsilon = 0.25$, as defined by the validation procedure in Section 4.6.1.

Table 5.2: Comparison of registration convergence success rates using different weighting schemes.

Weighting method	Success rate [%]
None	0.0
Binary mask	4.2
Distance-transform weighting	33.9

The results in Table 5.2 demonstrate that the standard unweighted approach consistently fails to provide reliable convergence in this environment, and hard binary masking alone is insufficient for robust inter-row registration. Completely removing bee-covered regions creates sharp discontinuities in the gradient space and significantly reduces the amount of usable structural information, leading to premature optimization failures.

In contrast, the proposed distance-transform-based weighting preserves spatially smooth contributions from regions near bees while effectively suppressing dynamic outliers. This substantially improves convergence robustness, increasing the registration success rate from 4.2% to 33.9%. While achieving convergence on roughly third of the pairs might initially seem limited, it is important to note that these inter-row pairs represent the most challenging alignments in the dataset, characterized by minimal spatial overlap and severe dynamic occlusions. Successfully recovering over 30% of these complex geometric constraints introduces more than enough redundancy into the global pose graph to effectively eliminate long-range odometry drift.

5.4 Tolerance to Bee Occlusion

The previous evaluation demonstrated the overall success rates of the distance-transform weighting scheme across different edge types. To comprehensively understand the boundary conditions and physical limits of the proposed

Weighted KLT algorithm, this experiment investigates how the localized density of bees directly influences the probability of successful convergence for both intra-row and inter-row transitions.

For each evaluated image pair in both the horizontal set ($\mathcal{P}_h$) and the vertical set ($\mathcal{P}_v$), an occlusion ratio was calculated. This ratio represents the percentage of pixels strictly within the valid overlapping region that are classified as bees by the binary detection mask (prior to applying the continuous distance transform). The image pairs were grouped into discrete bins based on their occlusion ratio, and the convergence success rate was evaluated independently for each bin.

Table 5.3: Registration success rate relative to the percentage of visual occlusion for **horizontal (intra-row)** image pairs $\mathcal{P}_h$.

Occlusion Ratio [%]	Number of Pairs	Success rate [%]
0-20	92	97.5
20-40	1696	98.5
40-60	3869	98.2
60-80	231	96.2
80-100	0	N/A

Table 5.4: Registration success rate relative to the percentage of visual occlusion for **vertical (inter-row)** image pairs $\mathcal{P}_v$.

Occlusion Ratio [%]	Number of Pairs	Success rate [%]
0-20	131	27.5
20-40	1191	43.0
40-60	4809	46.8
60-80	4371	17.9
80-100	72	4.2

The resulting distributions (see Figure 5.2) in Table 5.3 and Table 5.4 reveal a significant statistical and geometric disparity between the two edge types. For horizontal (intra-row) pairs, the spatial overlap between consecutive frames is massive (often exceeding 95% of the field of view). Consequently, extreme occlusion levels ($> 80\%$) are biologically rare, as they would require a dense, unbroken carpet of bees across almost the entire image. Due to this large structural redundancy, the Weighted KLT solver maintains a near-perfect success rate across all naturally occurring occlusion bins. Conversely, vertical (inter-row) pairs are restricted to a small overlap region, typically a narrow horizontal strip at the boundary of the image. This limited area acts as a spatial bottleneck. Even a moderately sized, localized cluster of bees residing

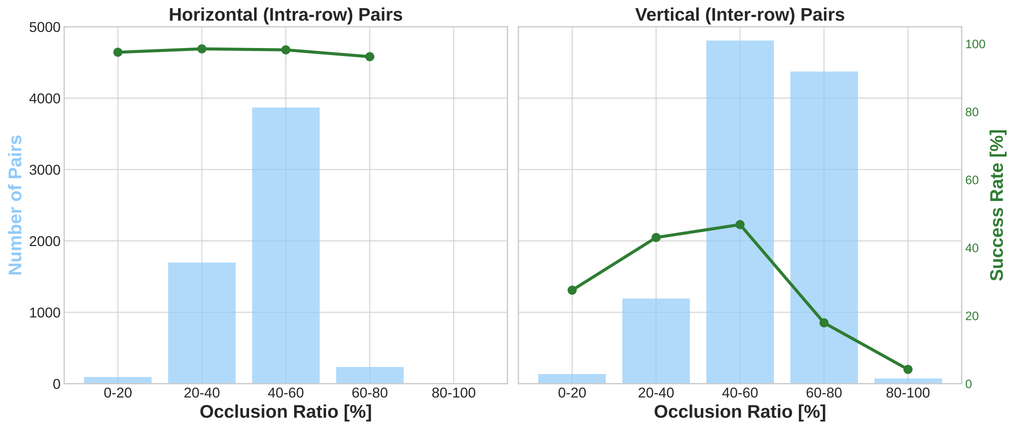
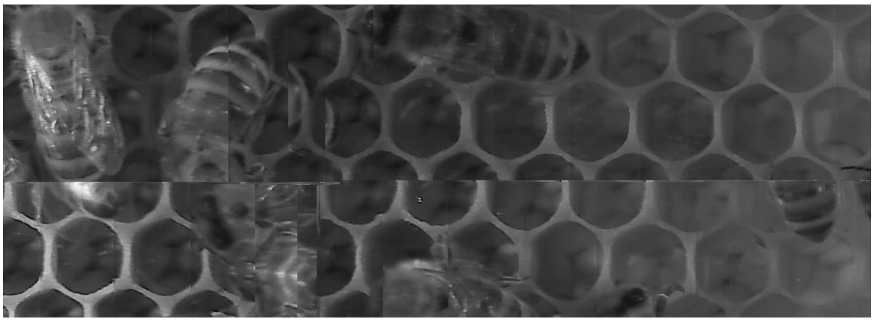


Figure 5.2: Comparison of registration success rates relative to the localized visual occlusion by bees. **(Left)** Horizontal (intra-row) pairs exhibit massive spatial overlap; consequently, extreme occlusion levels ($> 80\%$) are biologically rare, allowing the Weighted KLT to maintain robust convergence across all natural conditions. **(Right)** Vertical (inter-row) pairs are restricted to a narrow spatial bottleneck. The success rate sharply declines when localized occlusion exceeds 60%, as the overlapping region lacks sufficient static honeycomb structure to reliably constrain the affine parameters, leading to the necessary termination of the solver.

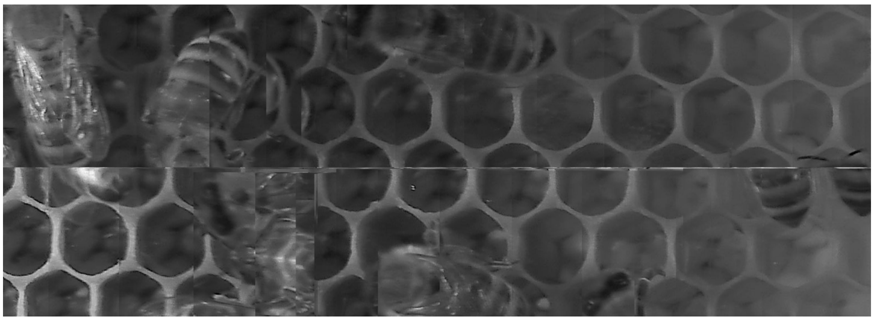
on this boundary can easily occlude the majority of the usable structural information. As observed in Table 5.4, the algorithm is relatively robust under low to moderate occlusion conditions ($< 60\%$), proving the efficiency of the dynamic weighting scheme. However, a sharp decline in the success rate occurs when the localized occlusion exceeds 60%. Under these conditions, the visual overlap region simply lacks sufficient static background (honeycomb edges) to constrain the affine parameters, leading to an ill-conditioned Hessian matrix and the necessary termination of the optimization. This analysis confirms that the overall failure rate on vertical edges is a direct consequence of the physical limits of visual registration in densely populated hive regions, rather than algorithmic instability.

5.5 Qualitative Evaluation

While the quantitative metrics presented in the previous sections provide an objective measure of the system's performance, a qualitative comparison of the resulting mosaics is essential to illustrate the impact of the proposed optimization pipeline. Figure 5.3 showcases a representative section of the honeycomb reconstructed using two different approaches. The initial map, constructed solely from raw mechanical odometry (see Fig. 5.3a), suffers from significant spatial drift and misalignment artifacts. Due to the accumulation of systematic errors during the serpentine scan, the cell boundaries appear blurred or duplicated (ghosting), making any subsequent semantic analysis or cell-state tracking impossible.



(a) : Initial map (raw odometry)



(b) : Optimized map (this work)

Figure 5.3: Qualitative comparison of the honeycomb reconstruction. (a) The map relying on mechanical odometry exhibits visible spatial drift. (b) The globally consistent mosaic achieved through the proposed optimization pipeline, showing sharp cell boundaries and accurate geometric alignment.

In contrast, the mosaic generated after applying the Weighted KLT alignment and global pose-graph optimization (see Fig. 5.3b) demonstrates a high degree of geometric consistency. The redundant inter-row constraints successfully eliminate the accumulated drift, resulting in sharp, well-defined cell structures that accurately represent the physical state of the comb. This visual improvement confirms that the proposed method provides a stable and reliable spatial foundation for future automated biological monitoring.

5.6 Robotic Platform Consistency Analysis

In addition to evaluating the quality of the final spatial reconstruction itself, the proposed framework also enables an indirect analysis of the acquisition platform's consistency. Since the visual registration estimates the relative displacement between neighboring frames independently of the robot odometry, the discrepancy between the odometric initialization and the visually estimated transformation can serve as a proxy indicator of the robotic system precision.

To investigate this, we analyze only intra-row image pairs, where the camera motion is approximately horizontal and where the image overlap is sufficiently large to ensure reliable visual registration. For each neighboring image pair, three metrics are evaluated:

1. **Odometry correction magnitude:** The absolute difference between the odometric displacement estimate and the displacement recovered by the KLT registration. The correction is evaluated separately for the horizontal and vertical components, as well as by its Euclidean norm.

2. **KLT convergence success rate:** The percentage of image pairs for which the weighted KLT optimization successfully converged below the predefined error threshold. In the context of intra-row pairs with high spatial overlap, a significant drop in convergence success is a strong indicator of hardware-level image degradation, such as motion blur from excessive vibrations, focus loss, or physical obstructions (e.g., propolis) on the observation glass.

3. **Affine deformation magnitude:** The deviation of the estimated affine component from the identity transformation. Since the robotic motion should ideally correspond to pure translation parallel to the hive plane, significant affine corrections may indicate mechanical instability, vibrations, or imperfect camera alignment.

These metrics provide an indirect estimate of the mechanical stability and calibration quality of the robotic acquisition system. Furthermore, systematic differences between individual robotic platforms or hive sides may reveal long-term degradation effects or the need for recalibration and maintenance.

Table 5.5: Robotic platform consistency and calibration metrics derived from intra-row visual alignment corrections. The analysis compares two robotic manipulators operating on opposite sides of Hive 0.

Data	Success Rate [%]	Transl. Error Median [px]	Transl. Error Mean [px]	Affine Dev. Median
Hive 0-A	94.2	3.0 ± 2.4	3.3	0.0022
Hive 0-B	96.0	6.9 ± 3.6	7.0	0.0021

As shown in Table 5.5, both robotic units exhibit exceptional mechanical consistency. The success rate of the visual registration remains above 94% for both sides, indicating a high quality of the captured images with sufficient structural overlap and minimal motion blur.

The median affine deviation is effectively zero (≈ 0.002) for both manipulators. This confirms a highly precise hardware calibration: the gantry axes are perfectly orthogonal, and the camera's optical axis remains strictly perpendicular to the observation glass throughout the scanning motion, with no detectable mechanical shear or rotational skew.

Regarding the translational component, the median errors reveal a minor, systematic mechanical drift of 3.0 px on Side A and 6.9 px on Side B per movement step. Notably, the mean and median values are nearly identical for both sides. This absence of significant heavy-tailed outliers suggests a smooth mechanical operation without abrupt hardware jerks or rail sticking.

While the robotic unit on Side B exhibits a slightly larger baseline drift than the unit on Side A, both values are extremely small relative to the image resolution.

Chapter 6

Discussion

This chapter provides an interpretation of the experimental results and evaluates the overall performance of the mapping pipeline. It examines the impact of the proposed weighting and regularization mechanisms on map integrity and acknowledges the limitations of the current implementation. Furthermore, it outlines several promising directions for future research.

6.1 Interpretation of Experiment Results

The experimental evaluation demonstrated that standard, unweighted registration approaches completely fail when applied to inter-row transitions in a populated observation hive. The introduction of the distance-transform-based weighting scheme proved to be the pivotal factor in achieving convergence, increasing the success rate to 33.9%.

While a failure rate of approximately 66% might initially appear high for standard computer vision tasks, it is a highly successful outcome within the specific context of our acquisition setup. Inter-row image pairs suffer from low spatial overlap, significant temporal delays (leading to shifting illumination), and dense, dynamic occlusions caused by bee movement. Successfully estimating over third of these extremely challenging constraints provides the global pose graph with more than enough redundancy to effectively eliminate the accumulated odometry drift.

The comparison between binary masking and distance-transform weighting highlights a fundamental property of gradient-based optimization. Hard binary masks introduce sharp, artificial discontinuities into the gradient space, which routinely trap the KLT tracker in local minima. By employing a continuous distance transform, the objective function landscape remains smooth, allowing the solver to gradually glide toward the global minimum while reducing the influence of photometric errors originating from the bees.

6.2 System Robustness and Regularization

The integrity of the final spatial map relies not only on the pairwise estimations but heavily on the formulation of the global optimization graph. Two specific regularization decisions were critical to the system's success:

- **Dynamic Huber Loss:** The use of Iteratively Reweighted Least Squares (IRLS) with a dynamically scaled Huber loss (via Median Absolute Deviation) proved essential. Even with the weighting scheme, some pairwise KLT registrations inevitably converge to incorrect local minima due to the self-similar hexagonal cells. The robust loss function effectively identifies these failures as outliers and down-weights their influence, preventing localized errors from corrupting the global map geometry.
- **Affine Regularization:** The unconstrained 6-DOF affine transformation model poses a significant risk in texture-less regions. Without constraints, the optimizer tends to artificially stretch or shear the image to minimize photometric error, leading to physically impossible honeycomb deformations. The introduction of the analytical Jacobian with the λ_{aff} regularization term successfully penalized these non-rigid components, forcing the system to rely on translation and rotation while allowing only micro-adjustments in scale and shear to compensate for minor camera perspective shifts.

6.3 Limitations of the Proposed System

Despite the robust geometric constraints, the current implementation has several inherent limitations that must be acknowledged:

1. **Dependence on Initial Odometry:** The Inverse Compositional KLT algorithm is a local optimization method. If the initial positional guess provided by the mechanical odometry drifts by more than half the width of a honeycomb cell (approximately 2.5 mm), the tracker is likely to converge to an adjacent, incorrect cell. The system cannot currently recover from such severe initial misalignments without external feature-matching mechanisms, which are largely ineffective in this environment.
2. **Complete Visual Occlusion:** In scenarios where the overlap region between two images is entirely covered by a dense layer of bees, the visual registration lacks sufficient structural information to compute a transformation. In such localized areas, the graph remains entirely dependent on the raw odometry.
3. **Computational Performance:** The pipeline was implemented in Python. While the use of the Inverse Compositional formulation and the `scipy.sparse` backend significantly accelerated the optimization compared to naive approaches, graph construction is still highly time consuming. Estimating transformations between vertical neighbors converges slowly mainly due to the shift in lighting.

6.4 Future Work

The methodology established in this thesis provides a solid foundation for possible future work:

- **Semantic Mapping:** With a globally consistent map available, future work can project the outputs of specialized biological object detectors (e.g., identifying eggs, larvae, pollen, and capped honey) directly onto the global coordinate frame. This would allow for automated, long-term spatial analysis of the colony's resource distribution and the queen's laying patterns.

- **Performance Optimization:** Porting the core KLT and Jacobian assembly routines to a compiled language such as C++ or utilizing libraries like Numba would significantly reduce execution time.

- **Integration of Queen-Tracking Data:** As the primary objective of the robotic platform is continuous queen monitoring, a vast dataset of queen-centered images is concurrently generated. Biologically, worker bees naturally maintain a spatial clearance around the queen (the retinue), which frequently exposes a clear view of the underlying honeycomb cells. Incorporating these unstructured, high-visibility frames into the pose graph could serve as robust geometric "bridges" between adjacent rows. This would significantly mitigate the effects of shifting illumination that currently complicate standard sequential inter-row registration and offer more spatial overlap, when connecting neighboring rows, making the algorithm more stable.

- **Multi-Scan Temporal Fusion:** The robotic observation platform routinely acquires multiple full-comb scans over the course of a single day. Because the spatial distribution of bees is highly dynamic, comb regions that are severely occluded in one scan may be entirely visible in a subsequent pass. Future work could leverage this temporal redundancy to synthesize a largely unoccluded global mosaic by systematically selecting the most static background pixels across multiple aligned scans. However, implementing this cross-scan fusion fundamentally requires the prior availability of accurate, globally consistent maps for each individual sequence, as accumulated inter-scan odometry drift must be fully compensated before pixel-level temporal fusion can occur.

- **Evaluated alternative: Geometric Semantic Classification:** While the Hough-transform-based geometric descriptor (see section 4.1) proved insufficiently robust for pairwise inter-frame registration, the underlying concept holds significant potential for post-processing semantic analysis. The physical geometry and size of a honeycomb cell directly dictate its biological function. For instance, the cell diameter acts as a physical stimulus for the queen's egg-fertilization reflex, determining worker versus drone castes, while the spatial dimensions of the cell have also been proven to actively influence epigenetic queen-worker differentiation [21]. Future work could adapt this geometric descriptor to operate on the completed, optimized global map. By applying explicit wall detection only to unoccluded and successfully reconstructed regions of the mosaic, the system could automatically classify cell types, extract structural anomalies, and provide a comprehensive spatial distribution of the comb's architecture.

- **Automated Diagnostic Reporting:** While the current mapping pipeline operates fully offline to conserve computational resources on the robotic hardware, this architecture is perfectly suited for batch-based hardware monitoring. Future work could integrate the proposed consistency metrics (affine deviation, translation error) into an automated post-scan diagnostic tool. After each completed full-comb scan, a server-side system could automatically process a sparse subset of intra-row pairs and generate a hardware health summary. This would allow operators to identify mechanical degradation, loose camera mounts, or calibration drift between scanning sessions.

Chapter 7

Conclusion

The objective of this bachelor thesis was to develop a robust system for the autonomous visual mapping of honeybee combs within an observation hive. To address this, we designed and implemented a complete image processing pipeline that successfully bridges the gap between raw robotic odometry and a globally consistent spatial map.

The core technical contribution lies in the implementation of a weighted Inverse Compositional KLT tracker, which, combined with an original distance-transform-based weighting, effectively suppresses the influence of bee movement. Furthermore, the global consistency of the map is ensured by a custom pose-graph optimization framework utilizing a robust Huber loss and analytical Jacobian formulations. Our experimental results confirmed that this approach significantly outperforms standard unweighted registration methods, achieving stable convergence even in the presence of severe occlusions and repetitive structural patterns.

This work establishes a reliable geometric foundation for future work. The generated spatially consistent maps serve as an essential prerequisite for future automated biological monitoring. By providing a stable coordinate frame, this system enables long-term tracking of individual cell states and large-scale analysis of colony dynamics, ultimately contributing to a deeper understanding of honeybee behavior and hive health.

Bibliography

[1] Martin Stefanec, Daniel N. Hofstadler, Tomáš Krajník, Ali Emre Turgut, Hande Alemdar, Barry Lennox, Erol Şahin, Farshad Arvin, and Thomas Schmickl. A minimally invasive approach towards "ecosystem hacking" with honeybees. *Frontiers in Robotics and AI*, 9, 2022.

[2] Jiří Ulrich, Martin Stefanec, Fatemeh Rekabi-Bana, Laurenz Alexander Fedotoff, Tomáš Rouček, Bilal Yağız Gündeğer, Mahmood Saadat, Jan Blaha, Jiří Janota, Daniel Nicolas Hofstadler, Kristina Žampachů, Erhan Ege Keyvan, Babür Erdem, Erol Şahin, Hande Alemdar, Ali Emre Turgut, Farshad Arvin, Thomas Schmickl, and Tomáš Krajník. Autonomous tracking of honey bee behaviors over long-term periods with cooperating robots. *Science Robotics*, 9(95):eadn6848, 2024.

[3] Randolf Menzel and Werner Backhaus. Color vision honey bees: Phenomena and physiological mechanisms. In Doekele Gerben Stavenga and Roger Clayton Hardie, editors, *Facets of Vision*, pages 281–297, Berlin, Heidelberg, 1989. Springer Berlin Heidelberg.

[4] Simon Baker and Iain Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[5] Rafael Barmak, Martin Stefanec, Daniel N. Hofstadler, Louis Piotet, Stefan Schönwetter-Fuchs-Schistek, Francesco Mondada, Thomas Schmickl, and Rob Mills. A robotic honeycomb for interaction with a honeybee colony. *Science Robotics*, 8(76):eadd7385, 2023.

[6] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[7] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 337–33712, 2018.

[8] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8922–8931, June 2021.

[9] Jiří Janota, Jan Blaha, Fatemeh Rekabi-Bana, Jiří Ulrich, Martin Stefanec, Laurenz Fedotoff, Farshad Arvin, Thomas Schmickl, and Tomáš Krajník. Towards robotic mapping of a honeybee comb. In *2024 International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, pages 1–6. IEEE, 2024.

[10] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 2005.

[11] M Nixon. Application of the hough transform to correct for linear variation of background illumination in images. *Pattern Recognition Letters*, 3(3):191–194, 1985.

[12] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, second edition, 2003.

[13] Katarzyna Bozek, Laetitia Hebert, Yoann Portugal, Alexander S. Mikheyev, and Greg J. Stephens. Markerless tracking of an entire honey bee colony. *Nature Communications*, 12(1):1733, 2021.

[14] Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[15] Paul W. Holland and Roy E. Welsch. Robust regression using Iteratively Reweighted Least-Squares. *Communications in Statistics - Theory and Methods*, 6(9):813–827, 1977.

[16] S.M. Pizer, R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, and K.E. Muller. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pages 337–345, 1990.

[17] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[18] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[19] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.

[20] Pauli Virtanen, R. Gommers, T.E. Oliphant, et al. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.

[21] Yuan Yuan Shi, Zachary Y. Huang, Zhi Jiang Zeng, Zi Long Wang, Xiao Bo Wu, and Wei Yu Yan. Diet and cell size both affect queen-worker differentiation through DNA methylation in honey bees (apis mellifera, apidae). *PLOS ONE*, 6(4):1–6, 04 2011.

Appendix A
Attachments

This appendix describes the structure of the electronic attachment submitted with this thesis.

- `src/` — Directory containing the complete Python source code.
  - `bee_blur.py` — Weights creation
  - `cells_detection.py` — *
  - `cells_match.py` — *
  - `config.py` — Stores constant variables (i.e images folder)
  - `graph_building.py` — Routines for processing data into graph structure.
  - `graph_opt.py` — Implementation of the pose-graph optimization.
  - `KLT.py`— Implementation of the Weighted KLT tracker.
  - `main.py` — executes the entire pipeline.
  - `mosaic.py` — Creation of mosaics using affine transformations or pure translations
  - results.py — Evaluation methods used in experiments.
  - `demo.ipynb` — Jupyter notebook with showcase of this work.
  - `bee_detect/` — Bee detection from Zdeněk Rozsypálek.
  - utls/
    - `cell.py` — *
    - `detection_utils.py` — *
    - `math_utils.py` — *
    - `file_utils.py` — Functions for loading and saving data.
    - `plot_utils.py` — Functions for results showcase.
    - structs.py — Structures used in this work.
- `data/` — Storage for images and bee detection model.
- `README.md` — Installation and usage instructions.
- archive*.zip — compressed dataset sample

* - LEGACY code used for approach 4.1, still kept for possible future use.