THE UNIVERSITY OF AUCKLAND
LIBRARIES AND LEARNING SERVICES

ResearchSpace@Auckland

Journal Article Version

This is the publisher's version. This version is defined in the NISO recommended practice RP-8-2008 http://www.niso.org/publications/rp/

Suggested Reference

Risse, B., Berh, D., Tao, J., Jiang, X., Klette, R., & Klaembt, C. (2013). Comparison of two 3D tracking paradigms for freely flying insects. *EURASIP Journal on Image and Video Processing*, *2013*. doi:10.1186/1687-5281-2013-57

Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

This is an open-access article distributed under the terms of the Creative Commons Attribution License

http://www.springeropen.com/about

http://www.sherpa.ac.uk/romeo/issn/1687-5176/

https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm

EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

Comparison of two 3D tracking paradigms for freely flying insects

Benjamin Risse[1,3*], Dimitri Berh[1], Junli Tao[2], Xiaoyi Jiang[1], Reinhard Klette[2] and Christian Klämbt[3]

Abstract

In this paper, we discuss and compare state-of-the-art 3D tracking paradigms for flying insects such as *Drosophila melanogaster*. If two cameras are employed to estimate the trajectories of these identical appearing objects, calculating stereo and temporal correspondences leads to an $\mathcal{NP}$-hard assignment problem. Currently, there are two different types of approaches discussed in the literature: probabilistic approaches and global correspondence selection approaches. Both have advantages and limitations in terms of accuracy and complexity. Here, we present algorithms for both paradigms. The probabilistic approach utilizes the Kalman filter for temporal tracking. The correspondence selection approach calculates the trajectories based on an overall cost function. Limitations of both approaches are addressed by integrating a third camera to verify consistency of the stereo pairings and to reduce the complexity of the global selection. Furthermore, a novel greedy optimization scheme is introduced for the correspondence selection approach. We compare both paradigms based on synthetic data with ground truth availability. Results show that the global selection is more accurate, while the previously proposed tracking-by-matching (probabilistic) approach is causal and feasible for longer tracking periods and very high target densities. We further demonstrate that our extended global selection scheme outperforms current correspondence selection approaches in tracking accuracy and tracking time.

Keywords: *Drosophila melanogaster*; Fruit flies; 3D tracking; Unscented Kalman filter; Global correspondence selection; Gibbs sampling; Greedy optimization

1 Introduction

The investigation of complex movement patterns of various organisms has become an integral subject of biological research. From a biological point of view, motion is the visual response to any kind of perceivable stimulation. The nervous system is responsible for the perception, the integration of the information, and the execution of the final response. One of the most popular model organisms to study how the nervous system controls locomotion is *Drosophila melanogaster* (i.e., fruit fly). Sophisticated genetic tools as well as advanced imaging techniques allow the functional dissection of neural circuits [1-4].

Drosophila is a holomethabolous insect. In the larval stage, locomotion is confined to two dimensions, whereas the adult fly moves in two and three dimensions. Approaches dealing with crawling larvae are common praxis; thus, two-dimensional (2D) tracking is well established in behavioral experiments [3,5-8]. In addition, flies confined to 2D motion are often used in behavioral experiments [2,9-11]. Basically, there are two ways to prevent the flies from takeoff: cutting the wings [12] or using an arena with a flat ceiling [13]. Both manipulations could lead to unnatural behavior [14]. Thus, three-dimensional (3D) tracking approaches are needed to address all kinds of behavioral phenotypes.

1.1 Related work

Work on freely flying fruit flies is still in its infancy, because it requires dynamic 3D correspondence analysis [15]. This analysis involves two challenging tasks: stereo matching (i.e., correspondence between camera views) and temporal tracking (i.e., correspondence over

*Correspondence: b.risse@wwu.de
[1]Department of Mathematics and Computer Science, University of Münster, Einsteinstraße 62, Münster 48149, Germany
[3]Department of Neuro and Behavioral Biology, University of Münster, Badestraße 9, 48149 Münster, Germany
Full list of author information is available at the end of the article

Springer

© 2013 Risse et al.; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

time). Together they form the so-called general multi-index assignment problem [16]. This problem is non-deterministically polynomial-time hard ($\mathcal{NP}$-hard) [16]. If all correspondences are known, triangulation is used to determine the 3D positions.

To avoid expensive multi-camera multi-target 3D tracking, existing approaches typically either track in two dimensions (no stereo matching) [2,9-11] or track only a single target (no ambiguities over time) [14,17]. If multi-camera multi-target 3D tracking is required, stereo matching and temporal tracking can be solved separately by accepting a decrease of tracking accuracy [18-20].

Among others, there are two fundamentally different paradigms used to capture 3D trajectories of multiple adult *Drosophila*. The first paradigm uses the extended Kalman filter and avoids complexity by separating stereo and temporal correspondence associations [21,22]. Due to this separation, optimal results cannot be guaranteed, and fragmented tracks prevent the preservation of the fly identities over time. The second paradigm performs a global selection by combining both tasks to calculate the overall best assignment [23]. As a result, identity preservation can be achieved for many flies and frames. However, the amount of possible combinations increases exponentially with the number of animals and time steps; thus, current solutions are only able to track for a short period.

Another probabilistic approach addresses the trade-off between identity preservation and long-term experiments [24]. The authors use the Hungarian algorithm and Kalman filtering for stereo matching and temporal correspondence association. Focusing on applicability for biologists, up to seven flies were evaluated in several experiments.

All the above-mentioned approaches focus on either tracking a few hundreds of targets for a short period of time or tracking less targets for more frames. High-density tracking is used in different research areas like particle tracking velocimetry [18,25] and tracking bats [26,27], bees [28] or fruit flies [20,23]. A quantitative comparison of several three-dimensional Lagrangian particle tracking approaches for high-density situations is given in [29].

Examples of long-term tracking approaches for fruit flies are given in [21,22,24,30,31]. In a recent publication, problems like noise and low frame rates are addressed to calculate trajectories of wild mosquitoes [32]. The authors used a probabilistic multi-target tracking for swarms of 6 to 25 mosquitoes. If hundreds of flies are tracked for a comparatively long period, trajectories are fragmented and the identity is not preserved. Furthermore, tracking several hundreds of flies simultaneously is not practical for most biological applications [24,30]. Only if swarming behavior needs to be analyzed, ambiguous animals are neglected leading to a strongly varying number of targets over time [33].

In a recent publication, multi-path branching was used to handle occlusions by employing global optimization when calculating the trajectories [34]. The algorithm was exhaustively tested for both high-density and long-term situations. Again, the tracking accuracy decreases if the number of targets and the number of frames increases simultaneously.

1.2 Proposed algorithms and comparison scheme

In this paper, we compare identity preserving 3D tracking approaches for long-term experiments considering biological usability. First, we present algorithms for both the above-mentioned paradigms (see Figure 1):

- The previously proposed tracking-by-matching (TbM) solution [35] integrates a third camera to conduct projection consistency check into the probabilistic approach.
- In addition, we introduce a global correspondence selection (GCS) algorithm (extension of [23]), calculating the global search space and minimizing a cost function afterwards.

Limitations of the TbM and the GCS approach are addressed by utilizing a third camera to verify the consistency of stereo pairings. The third camera is integrated by the so-called projection consistency [35]. As a result, the amount of ambiguous temporal associations is reduced in the TbM approach. The GCS approach benefits from the projection consistency by means of a reduced overall complexity. Besides utilizing Gibbs sampling for optimization, as suggested by [23], we introduce an alternative greedy selection scheme (see Figure 1). It should be pointed out that we use GCS in terms of optimizing a global search space, not determining the global optimum for our optimization task.

We compare both paradigms, the TbM and the GCS approach, based on synthetic data; thus, the ground truth is available. Global correspondence selection was done via Gibbs sampling [23] and greedy optimization utilizing projection consistency. This leads to the comparison scheme illustrated in Figure 1.

This paper is organized as follows: In Section 2, we provide notes about notations and central equations. In particular, the projection consistency is described in detail. Algorithms are presented in Section 2.2. Section 2.2.3 describes the extensions of the GCS approach. The synthetic data and measures used for comparison are described in Section 3. All results are listed in Section 4: We compare GCS approaches with the TbM approach in Sections 4.2, 4.3, and 4.4. In addition, we compare both GCS approaches in more detail in Section 4.1. A concluding discussion of both paradigms is given in Section 5.

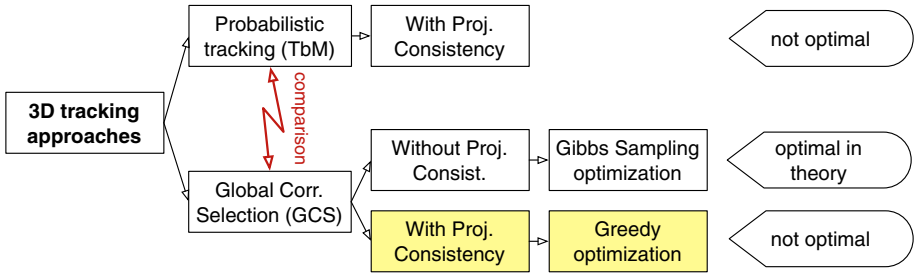

Figure 1 General comparison scheme of this paper. The new approach is highlighted in yellow.

2 Methods

Both algorithms expect time-synchronized image streams from up to three cameras. Let $\mathcal{I}_t^i$ represent these images of cameras $i = 1, 2, 3$, and time $t = 1, \ldots, T$. All cameras need to be calibrated; thus, the camera matrices $\mathbf{K}^i$, rotation matrices (from camera i to camera j) $\mathbf{R}^{ij}$, and translation vectors $\mathbf{t}^{ij}$ are given. Then, the fundamental matrices can be calculated by

$$\mathbf{F}^{ij} = \left(\mathbf{K}^j \mathbf{t}^{ij}\right) \times \left(\mathbf{K}^j \mathbf{R}^{ij} (\mathbf{K}^i)^{-1}\right) \tag{1}$$

(for more details, see [36]). Consider a swarm of flying targets of similar appearance and small size. The centers of detected targets (i.e., blobs) in a single image $\mathcal{I}_t^i$ are denoted by $M_t^i = \{m_{n^i,t}^i\} = \{(x, y)\}$ for $n^i = 1, \ldots, N_t^i$ targets at time t, where (x, y) is the image coordinate of the objects' centroid. The value N_t^i may differ due to occlusions or noise.

To calculate the 3D positions of the flies, stereo correspondences between detected blobs need to be established. Since we use three cameras, triplets of image points $(m_{n^1,t}^1, m_{n^2,t}^2, m_{n^3,t}^3)$ correspond to one target. In general, two 2D image coordinates are sufficient to calculate a single 3D position; thus, we define all possible pairs given by $H_t^{ij} = M_t^i \times M_t^j$ between camera i and j. A pairing $s_{k,t}^{ij} \in H_t^{ij}$ could represent either a true or a false correspondence for target k.

2.1 Stereo matching and projection consistency

Both paradigms perform stereo matching based on epipolar geometry and verify matches using the so-called projection consistency constraint [35].

2.1.1 *Stereo matching*

Stereo matching is used to identify possible pairings between two respective views and thus result in possible correspondences. For matching a point $m_{n^i,t}^i$ in $\mathcal{I}_t^i$ with a point $m_{n^j,t}^j$ in $\mathcal{I}_t^j$, both need to be located on the same epipolar line. Given $m_{n^i,t}^i$, the corresponding epipolar line $\mathbf{l}_{n^i}^j$ in $\mathcal{I}_t^j$ can be calculated by

$$\mathbf{l}_{n^i}^j = \mathbf{F}^{ij} m_{n^i,t}^i. \tag{2}$$

Detected points from M_t^j lying on this epipolar line are matched to $m_{n^i,t}^i$ and indicate possible pairings $\{s_{k,t}^{ij}\} \subset H_t^{ij}$.

2.1.2 *Projection consistency*

Since we use three calibrated cameras, triplets of 2D points $(m_{n^1,t}^1, m_{n^2,t}^2, m_{n^3,t}^3)$ located in images $\mathcal{I}_t^1, \mathcal{I}_t^2$, and $\mathcal{I}_t^3$ correspond to the same target in the 3D space. Projection consistency is applied to those triplets to verify the overall match. This constraint is satisfied if the respective projections from two 2D points $(m_{n^i,t}^i, m_{n^j,t}^j)$ into the third view $\mathcal{I}_t^h$ are *sufficiently close* to $m_{n^h,t}^h$. To project $(m_{n^i,t}^i, m_{n^j,t}^j)$ into $\mathcal{I}_t^h$, we calculate the 3D coordinate p_t^h by triangulating $(m_{n^i,t}^i, m_{n^j,t}^j)$ and use the camera matrices from $\mathcal{I}^h$ to obtain the hypothetical 2D position $\widetilde{m}_t^h$ in $\mathcal{I}_t^h$. Afterwards, we search for the closest point $m_{*,t}^h$ in M_t^h by

$$m_{*,t}^h = \min_{n^h \in \{1,\ldots,N_t^h\}} \left(\mathrm{dist}(\widetilde{m}_t^h, m_{n^h,t}^h) + (\mathbf{l}_{n^i}^h)^T m_{n^h,t}^h \right), \tag{3}$$

where the first summand is the Euclidean distance between the hypothetic position $\widetilde{m}_t^h$ and the measured positions in M_t^h and the second summand is the distance between a measured point and the epipolar line $\mathbf{l}_{n^i}^h$ in view $\mathcal{I}^h$ corresponding to view $\mathcal{I}^i$. If

$$\mathrm{dist}(\widetilde{m}_t^h, m_{*,t}^h) < \tau, \tag{4}$$

then blob $\widetilde{m}_t^h$ (and thus the underlying pairing $(m_{n^i,t}^i, m_{n^j,t}^j)$) describes a correct stereo correspondence. τ is the threshold for the projection consistency and depends on calibration accuracy. The triplet $(m_{n^i,t}^i, m_{n^j,t}^j, m_{n^h,t}^h)$ satisfies the projection consistency if $m_{n^h,t}^h = m_{*,t}^h$, for all possible combinations $i, j, h \in \{1, 2, 3\}$, $(i \neq j \neq h)$.

2.2 Presented algorithms

To compare current state-of-the-art tracking paradigms, we introduce the TbM algorithm and a GCS algorithm.

We try to overcome the limitations of probabilistic tracking, namely the separation of stereo matching and temporal tracking, by integrating projection consistency into the temporal tracking routine. Exponential complexity, arising in global correspondence selection algorithms, is avoided by reducing the global search space based on the projection consistency. Correspondence selection can be done by Gibbs sampling [37] or in a greedy manner. Since we introduce a novel selection scheme for the GCS approach (yellow box in Figure 1), we describe it in more detail. The TbM is explicitly described in [35].

2.2.1 Tracking-by-matching algorithm

As in all probabilistic tracking approaches, our tracking algorithm models the position and motion information of the targets independently from the stereo correspondences between the views. We use the *unscented Kalman filter* (UKF) as a Bayesian framework for 2D tracking [8,38]. Using the notation introduced above, every target $m^i_{n^i,t}$ in every view ($i \in \{1,2,3\}$) is represented by its own tracker T^i_k (i.e., a single UKF). Temporal tracking is achieved by referring one of the measured n^i-th targets to a specific tracker k over time t (yellow box 'temporal tracking' in Figure 2). Thus, for each new frame triplet, every UKF predicts the next possible 2D position for its target. Then, detections close to the predictions are verified with projection consistency (green box 'Verify new triplet via proj. consist.'; for details, refer to [35]). In this way, the projection consistency constraint is used to integrate stereo matching into temporal tracking. After updating all trackers T^i_k, this procedure is repeated as long as there are further frames available (see Figure 2).

2.2.2 Global correspondence selection

Before going into formal details, the following section introduces the GCS algorithm in a top-down manner.

General workflow of the GCS approach. In distinction from the TbM approach, temporal tracking and stereo matching is not done within the main loop (compare to Figure 2). In fact, a global search space named $\mathfrak{S}$ is constructed over all the accessible frames before the actual tracking (compare to box 'Construction of $\mathfrak{S}$'). Afterwards, the best possible assignments are calculated by minimizing an overall cost function operating on $\mathfrak{S}$.

To reduce the size of this search space, epipolar and temporal assumptions are made before considering actual correspondences. As illustrated in Figure 2, possible stereo correspondences between cameras 1 and 2 are calculated for every time step. Only blob pairings close to their respective epipolar lines are considered as possible stereo matches (box 'Calculate stereo correspondences'). If projection consistency is used (dashed box in Figure 2), invalid matches are removed or replaced via Equation 4. Note that the third camera is only used to replace incorrect matches from cameras 1 and 2. In other words, projection consistency is used to further reduce the set of possible pairings arising from $\mathcal{I}^1$ and $\mathcal{I}^2$. The resulting set of matches can be interpreted as a set of possible 3D positions. Given two sets of 3D positions for consecutive time steps, possible temporal assignments can be calculated (compared to box 'Calculate temporal correspondences'). If a target has no successor within a 3D neighborhood (given by the maximal flight speed), it is removed from $\mathfrak{S}$.

This reduction is done for all available frames and time steps T. After constructing the search space, several assignments are unique. Ambiguous pairings and ambiguous temporal correspondences form natural clusters in $\mathfrak{S}$. Thus, only samples inside these clusters need to be optimized (see box 'Get ambiguous clusters from $\mathfrak{S}$' in Figure 2). The subsequent optimization is done by a cost function introduced below which incorporates stereo and temporal matching (Equation 6). We implemented two optimization strategies to find possible samples in the respective clusters (see box 'Greedy / Gibbs cluster optimization').

Formal definition of the search space. Let $\mathbb{P}(H^{ij}_t)$ be the power set of all pairings $s^{ij}_{k,t} \in H^{ij}_t$. To avoid additional complexity, arising from pairwise pairings between three views, we use cameras 1 and 2 for stereo matching. Thus, the initial search space is constructed for H^{12}_t. Blobs from $\mathcal{I}^3_t$ are only considered if projection consistency is used (see Section 2.2.3).

The subset containing N_t pairings from the power set $\mathbb{P}(H^{12}_t)$ is given by the set of N_t permutations $\mathcal{P}_{N_t}(H^{12}_t) \subset \mathbb{P}(H^{12}_t)$ (i.e., $\mathcal{P}_{N_t}(H^{12}_t)$ contains all possible combinations of pairings with N_t elements). A single set $S^{12}_t = (s^{12}_{1,t}, s^{12}_{2,t}, \ldots, s^{12}_{N_t,t}) \in \mathcal{P}_{N_t}(H^{12}_t)$ is called a configuration and contains N_t stereo correspondences for time step t. If camera indices are not necessary, we use $(s_{1,t}, s_{2,t}, \ldots, s_{N_t,t}) = S_t = S^{ij}_t$ and $H_t = H^{ij}_t$ for $i,j \in \{1,2,3\}$, $i \neq j$.

Let $\mathfrak{S} = (\mathbb{P}(H_1), \mathbb{P}(H_2), \ldots, \mathbb{P}(H_T))$ be the set of all configurations over all time steps, or $\mathfrak{S}_{\mathcal{P}} = (\mathcal{P}_{N_1}(H_1), \mathcal{P}_{N_2}(H_2), \ldots, \mathcal{P}_{N_T}(H_T))$ if the number of targets per time step is known. A sequence of configurations between two time steps $t-1$ and t is denoted by $S_{t-1:t}$ and contains temporal correspondences between consecutive frames. Thus, an overall solution, containing all tracks for all flies and T time steps, is given by a sequence $S_{1:T} \in \mathfrak{S}$. The entire 3D trajectory of target k is then given by $s_{k,1:T} = (s_{k,1}, s_{k,2}, \ldots, s_{k,T})$.

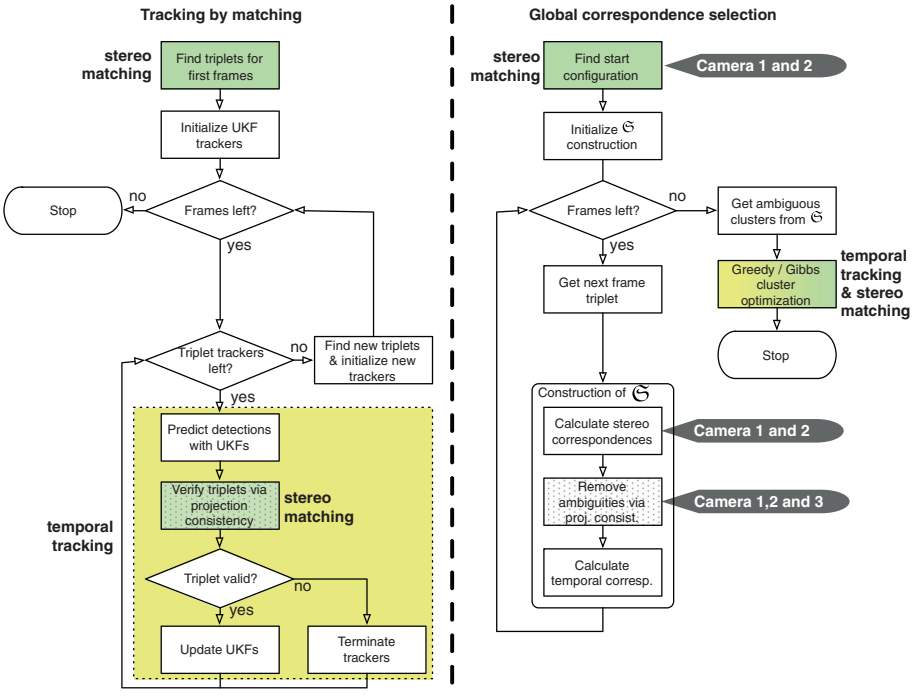

Figure 2 Flow charts of the two 3D tracking paradigms. Temporal tracking is marked in yellow, stereo matching is marked in green, and projection consistency is highlighted by dots.

Cost function. Stereo matching and temporal tracking is incorporated into a single optimization task, solving the optimization problem

$$S^*_{1:T} = \arg\min_{S_{1:T}\in\mathfrak{S}} f(S_{1:T}). \tag{5}$$

The cost function $f(\cdot)$ incorporates an epipolar constraint $f_E(\cdot)$ for stereo matching, kinetic coherence $f_K(\cdot)$ for temporal tracking, and a so-called conservation-observation match $f_C(\cdot)$ to punish multiple assignments. Thus, $f(\cdot)$ can be written as a sum of all the abovementioned constraints

$$f(S_{1:T}) = \alpha \sum_{t=1}^{T} f_E(S_t) + \beta \sum_{t=1}^{T} f_C(S_t, H_t) + \gamma \sum_{t=1}^{T} f_K(S_{t-1}, S_t) \tag{6}$$

with weights α, β, and γ (compare to [23]).

Cost function summands. Epipolar costs are defined as

$$f_E(S_t) = \sum_{k=1}^{N_t} \rho_e(s_{k,t}),$$

where $\rho_e(s_{k,t})$ sums the distances between the blobs $m^i_{k,t}, m^j_{k,t}$ from $s_{k,t}$ to its epipolar lines (compare Section 2.1.1). To avoid improbable stereo matchings, values $f_E(\cdot)$ larger than a threshold ε_E are set to ∞:

$$f_E(S_t) = \begin{cases} \sum_{k=1}^{N_t} \rho_e(s_{k,t}) & \forall k:\ \rho_e(s_{k,t}) < \varepsilon_E \\ \infty & \text{otherwise.} \end{cases} \tag{7}$$

The kinetic coherence

$$f_K(S_{t-1}, S_t) = \sum_{k=1}^{N_t} \rho_k(s_{k,t-1}, s_{k,t}) = \sum_{k=1}^{N_t} \mathrm{dist}(p_{k,t-1}, p_{k,t})$$

calculates the Euclidean distances dist$(\cdot)$ between 3D positions $p_{k,t-1}$ and $p_{k,t}$ (defined by $s_{k,t-1}$ and $s_{k,t}$). $\rho_k(\cdot)$ expects two consecutive pairings for 3D coordinate calculation. Improbable temporal connections are set to ∞:

$$f_K(S_{t-1}, S_t) = \begin{cases} \sum_{k=1}^{N_t} \rho_k(s_{k,t-1}, s_{k,t}) & \forall k : \rho_k(s_{k,t-1}, s_{k,t}) < \varepsilon_K \\ \infty & \text{otherwise.} \end{cases} \tag{8}$$

Finally, the conservation observation match is defined as

$$f_C(S_t, H_t) = \frac{1}{N_t^1} \sum_{k=1}^{N_t^1} |n_c(m_{k,t}^1, S_t) - 1| + \frac{1}{N_t^2} \sum_{k=1}^{N_t^2} |n_c(m_{k,t}^2, S_t) - 1|,$$

where $n_c(m_{k,t}^i, S_t)$ adds up the contributions of a blob $m_{k,t}^i$ in configuration S_t. If the number of correspondences exceeds a threshold ε_C, configuration costs are set to ∞:

$$f_C(S_t, H_t) = \begin{cases} \rho_c(S_t, M_t^1) + \rho_c(S_t, M_t^2) & \forall k : n_c(m_{k,t}, S_t) < \varepsilon_C \\ \infty & \text{otherwise.} \end{cases} \tag{9}$$

where $\rho_c(S_t, M_t^i) = \frac{1}{N_t^i} \sum_{k=1}^{N_t^i} |n_c(m_{k,t}^i, S_t) - 1|$.

Recursive decomposition. Equation 6 can be rewritten in a recursive manner as follows:

$$f(S_{1:T}) = f(S_{1:t-1}) + \Delta f(S_t) \tag{10}$$

with

$$\Delta f(S_t) = \alpha f_E(S_t) + \beta f_C(S_t, H_t) + \gamma f_K(S_{t-1}, S_t). \tag{11}$$

Thus, the whole optimization can be done by dynamic programming (for more details, see [23]).

Reduction of $\mathfrak{S}$. In [23], Gibbs sampling [37] is suggested to find the best possible sequence of configurations $S_{1:T}^* \in \mathfrak{S}$. Since $\mathfrak{S}$ is a set of T power sets $\mathbb{P}(H_t)$, several steps are suggested to reduce the search space. First of all, sampling for solutions with N_t targets for time t leads to a reduced set $\mathcal{P}_{N_t}(H_t)$. Thus, we redefine the overall search space for $S_{1:T}^*$ by $\mathfrak{S}_\mathcal{P} = (\mathcal{P}_{N_1}(H_1), \mathcal{P}_{N_2}(H_2), \ldots, \mathcal{P}_{N_T}(H_T))$.

The set H_t is reduced by rejecting pairings which do not satisfy Equation 7. In the remaining subset $\mathfrak{H}_t \subset H_t$, only blob pairings $s_{k,t}$ close to the respective epipolar lines are considered. Due to the recursive decomposition given in Equation 10, the successor to S_{t-1} can be selected from the N_t permutation $\mathcal{P}_{N_t}(\mathfrak{H}_t)$. Since kinetic costs are limited (see Equation 8), improbable temporal correspondences can be rejected from $\mathcal{P}_{N_t}(\mathfrak{H}_t)$. Figure 3 illustrates the reduction of the cardinality for an N_1 permutation.

After rejecting both impossible stereo matchings and temporal correspondences, some sequences of configurations $S_{t-1:t} \in (\mathcal{P}_{N_{t-1}}(\mathfrak{H}_{t-1}), \mathcal{P}_{N_t}(\mathfrak{H}_t))$ are unique. The remaining ambiguities form natural clusters $\mathcal{C}_{(t-\delta:t),\nu} \subset \mathfrak{S}_\mathcal{P}$ for $\delta + 1$ frames and ν flies. Zou et al. [23] extend ambiguous clusters by adjacent pairings. However, these pairings can again be involved in an ambiguous cluster. Since we tried to keep the identity over time, we merged the clusters in these situations as long as there are no ambiguous situations before and after each cluster anymore. In this way, the resultant clusters include overall ambiguous situations, and the domain of Equation 5 is global.

Since the cluster size increases exponentially with the number of targets N and time steps T, Gibbs sampling also requires thousands of sampling steps to guarantee good results. Indeed, the authors of [23] were only able to track for less than 1 s of recording.

2.2.3 Introduced improvements for the GCS approach

Here, we introduce two extensions to improve the performance of the GCS approach:

- Utilizing projection consistency to reject ambiguous pairings $s_{k,t}$ and thus reducing the sizes of the clusters
- Performing optimization in a greedy manner by selecting the best successor directly based on Equation 11

GCS with projection consistency. Similar to the above introduced probabilistic tracking approach, ambiguities and wrong stereo matches increase the size of the search space $\mathfrak{H}_t^{12}$. Thus, all pairings $s_{k,t}^{12} \in \mathfrak{H}_t^{12}$ are projected into the view of the third camera $\mathcal{I}^3$. Only pairings satisfying Equation 4 remain in $\mathfrak{H}_t^{12}$; inconsistent pairings are rejected.

An ambiguity is found if two pairings $s_{i,t}^{12}, s_{j,t}^{12}$ in the remaining $\mathfrak{H}_t^{12}$ share a single blob $m_{k,t}$. The blob $m_{*,t}^3$ satisfying Equation 4 is then used to generate a new pairing $s_{i,t}^*$, containing $m_{*,t}^3$ from $\mathcal{I}_t^3$ and the unambiguous blob from $\mathcal{I}_t^1$ or $\mathcal{I}_t^2$. Afterwards, ambiguous pairings in $\mathfrak{H}_t^{12}$ are replaced by unique pairings $s_{i,t}^*$.

Let $\mathfrak{H}_t^*$ be the reduced subset, containing the new unique pairing $s_{i,t}^*$ (compare '$\mathfrak{H}^*$ with PC' in Figure 4). The overall search space for Equation 5 is then given by $\mathfrak{S}_\mathcal{P}^* = (\mathcal{P}_{N_1}(\mathfrak{H}_1^*), \mathcal{P}_{N_2}(\mathfrak{H}_2^*), \ldots, \mathcal{P}_{N_T}(\mathfrak{H}_T^*))$.

The optimization of clusters based on Equation 5 via Gibbs sampling is described in [23]. The greedy optimization strategy is described below.

Greedy optimization. Given a cluster with ambiguities $\mathcal{C}_{(t-\delta:t),\nu} \subset \mathfrak{S}_\mathcal{P}^*$, a sequence of configurations $S_{t-\delta:t} \in$

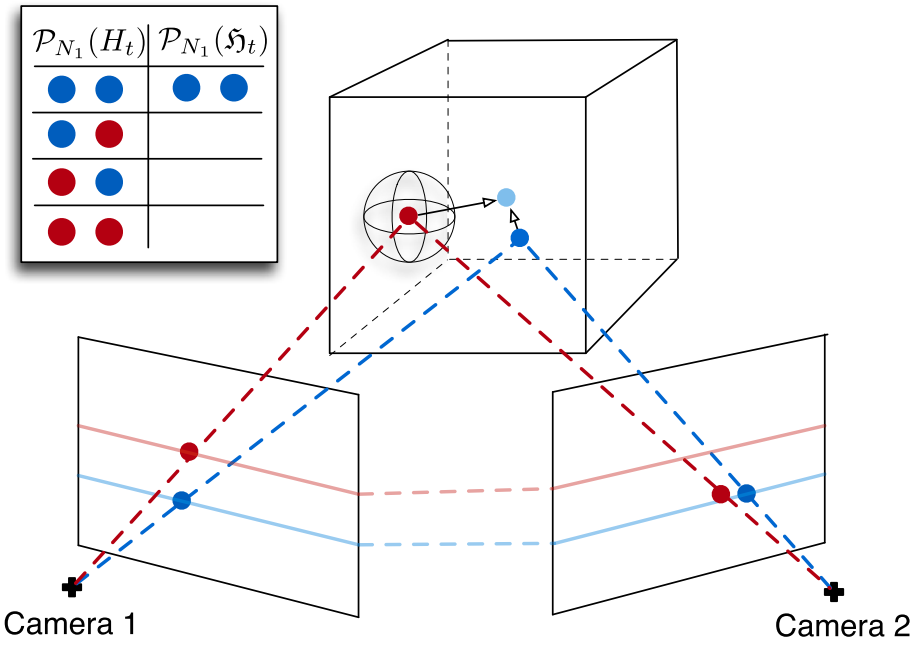

Figure 3 Example for the cardinality reduction of $\mathfrak{S}$. For a given time t, only one target needs to be found; thus, $\mathbb{P}(H_t)$ reduces to $\mathcal{P}_{N_1}(H_t)$. The resultant search space is illustrated in column $\mathcal{P}_{N_1}(H_t)$, containing only combinations of two blobs. The red blob detections do not satisfy the epipolar constraint (illustrated via red epipolar lines). In addition, the kinetic costs exceed the limit (indicated by the search sphere around the red blob). Thus, the cardinality of $\mathcal{P}_{N_1}(H_t)$ further decreases to $\mathcal{P}_{N_1}(\mathfrak{H}_t)$, containing a single unique pairing.

$\mathcal{C}_{(t-\delta:t),\nu}$ must be determined. Let the sequence of configurations be $S_{t-\delta:t} = (S_0, S_1, \ldots, S_\delta)$, which must be optimized for ν flies. For a given $S_i = (s_{1,i}, s_{2,i}, \ldots, s_{\nu,i})$, a successor to every pairing $s_{k,i}$ is selected based on Equation 11, by choosing the pairing $s_{*,i+1}$ with minimal costs (starting with $i = 0$ and $k = 1$). If, for example, $s_{1,0}$ is already assigned to $s_{1,1}$, a successor to $s_{2,0}$ is selected by $s^*_{2,1} = \arg\min_{s_{2,1} \in S_1} \alpha \sum_{i=1}^{N_1} f_E(S_1) + \beta \sum_{i=1}^{N_1} f_C(S_1, H_1) + \gamma \sum_{i=1}^{N_1} f_K(S_0, S_1)$. This is successively done for all pairings and all configurations until every pairing in every configuration has a successor.

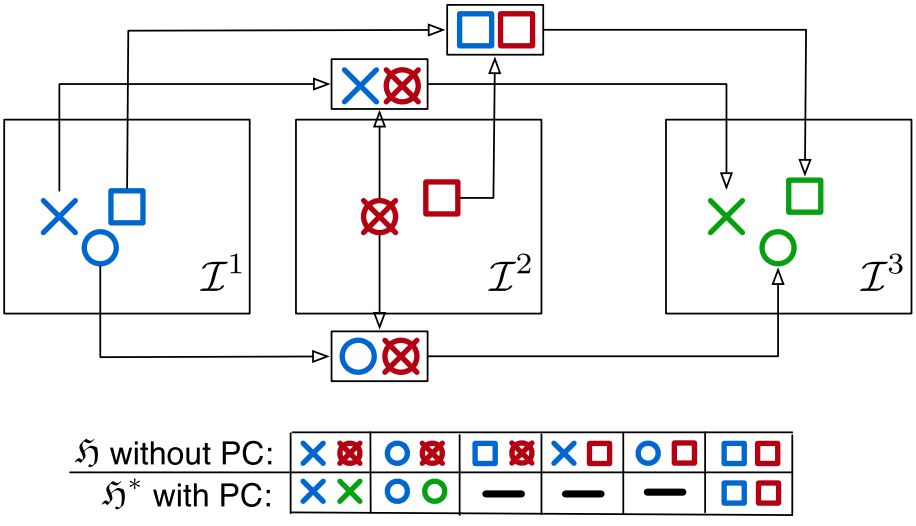

Figure 4 Example for pairings $\mathfrak{H}$ with and without projection consistency (PC). One target is occluded in $\mathcal{I}^2$, and all possible combinations are generated between $\mathcal{I}^1$ and $\mathcal{I}^2$. Pairings that do not satisfy the PC constraint are removed, and ambiguous pairings are corrected using projection consistency.

2.2.4 Complexity of the algorithms

The complexity of the GCS search space and thus the memory storage is $\mathcal{O}(k^{NT})$ in theory (N is the number of targets, T is the number of time steps, and $k \leq N$ denotes ambiguities after cardinality reduction), since there are k^N possible configurations between two views and each of these configurations at t can be combined with all configurations at $(t+1)$. Optimization is only necessary for ambiguous clusters $\mathcal{C}_{(t-\delta:t),\nu}$, therefore $N = \nu$ specifies the number of flies in this cluster and $T = \delta + 1$ specifies the length of the cluster. Thus, the global optimum must be calculated based on k^{NT} possible cluster configurations. Given the recursive decomposition of Equation 10, the complexity of the cost calculations decreases to $\mathcal{O}(Tk^N)$. The theoretical computational complexity is $\mathcal{O}(NTL)$ for Gibbs sampling (where L is the number of sampling steps) and $\mathcal{O}(NT)$ for greedy optimization.

For TbM algorithm, no extra space for tracking is needed. Thus, the memory storage complexity is $\mathcal{O}(NT)$. The time complexity is $\mathcal{O}((NT)^2)$. For the first three frames, exhausted search for triplets requires matching all detections in two views.

3 Experiments

3.1 Synthetic data

Both tracking paradigms are evaluated using synthetic data, generated by the swarm simulator introduced in [35]. The simulator generates all necessary data for tracking (i.e., rendered images and camera matrices) and evaluation (i.e., ground truth of the 2D and 3D trajectories). For our tests, three synchronized and calibrated cameras are placed around a $20 \times 20 \times 20$ cm^3 chamber. All movies are recorded with 800×800 pixel resolution and 150 fps. Since the beam width of the field of view is 45°, all cameras are placed 80 cm away from the cube's center. Rotations around the y-axes, for cameras 1, 2, and 3, are 0°, -120°, and 120°, respectively.

According to [39], the maximum flight speed is set to 0.8 m/s. The crawling speed is reduced by the factor 0.1, and we use a Gaussian random walk for flight movement calculation [35]. To achieve more realistic conditions and to increase the probability of occlusions and nearby targets, we integrated *negative geotaxis* within our random walk model. Negative geotaxis describes the tendency of *Drosophila* to orient themselves against the earth's gravity [40]. We integrated negative geotaxis by manipulating the randomly generated velocity in the y direction $v_t = \Theta v_{t-1} + n_t$ (with Gaussian noise $n_t \in \mathcal{N}(0, \sigma^2)$ and smoothness $\Theta \in [0, 1]$). With a probability of 0.002% the y entry of n_t is forced to be zero or positive over time.

In this way, we generated several test movies with an increasing number of targets. For most real-world locomotion experiments, 50 flies per run are sufficient; thus, we generated movies with 10 to 50 targets and 1,000 frames (approximately 6 s; Sections 4.2 and 4.1). In addition, we made a long-term movie with 50 flies over 3,000 frames (Section 4.3) and high-density movies with a few hundreds of flies and time steps (Section 4.4).

To guarantee identical raw data for both algorithms, the 2D positions of all views are established by a separate blob detection routine. Resultant measurements contain time steps with several occluded flies in all views (leading to changes in N_t^i; compare to Table 1). We also added Gaussian noise ($\sigma^2 = 0.001$ in the intensity domain $[0, 1]$) to the ground truth videos to simulate blob detections under realistic conditions. Figure 5 shows an example triplet of noisy images of 200 flies.

3.2 Evaluation and comparison measure

Both paradigms are compared in terms of tracking accuracy using the *correspondence and association errors (E_{ca})* [35]. The E_{ca} is defined as follows:

$$E_{\text{ca}} = \frac{N_c + N_a}{T}, \tag{12}$$

where N_c is the number of incorrect stereo matches, N_a is the number of false temporal associations, and T is the number of frames. To calculate N_a and N_c, all computed 3D trajectories are assigned to their respective ground truth paths. This assignment is used to calculate Euclidean distances between calculated positions and ground truth positions. If the distance is not within a tolerance, N_c is incremented for each frame and time step. The temporal association value is incremented if the ID of the calculated 3D paths changes between consecutive frames.

4 Results

We tested all combinations illustrated in Figure 1 as follows:

- Tracking by matching method (named *TbM*)
- GCS optimized via Gibbs sampling analogous to [23] (named *Gibbs*)
- GCS with projection consistency (PC) optimized via Greedy (named *Greedy PC*)

General tracking results for 50 flies and over 1,000 time steps are given in Figure 6.

Table 1 summarizes results for all approaches. The resultant E_{ca} value is additionally plotted in Figure 7a.

4.1 Gibbs sampling vs. greedy optimization

The first observation is related to the number of occlusions and maximal cluster sizes. In general, the complexity of the global search space increases with the number of targets and frames [16]. If there are only a few ambiguities (e.g., occlusions, nearby 3D paths), most of the corre-

Table 1 Overview of the results

		General					Long term	High density		
(N, T)		(10,1000)	(20,1000)	(30,1000)	(40,1000)	(50,1000)	(50,3000)	(100,200)	(150,150)	(200,150)
Occlusions	Absolute	21	119	289	846	1248	3,295	889	1,850	3,423
	Relative	0.007	0.04	0.096	0.282	0.416	0.366	1.482	4.111	7.607
max $\lvert\mathcal{C}\rvert$	Gibbs	102	19,024	1,629	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$	$\lvert\mathfrak{S}_{\mathcal{P}}\rvert$
	Greedy PC	94	2,140	194	516	713	1,928	16,343	9,946	13,413
Time	TbM	27.68	38.22	48.26	65.61	78.63	170.4	34.07	50.59	71.0
	Gibbs	93.7	3,828.6	1,069.9	530.8	712.8	>4 h	>4 h	>4 h	>4 h
	Greedy PC	71.1	87.6	110.8	160.2	200.6	748.4	2,560.49	401.56	3,114.2
E_{ca}	TbM	0.009	0.052	0.33	0.326	0.544	0.591	1.16	3.3533	6.7267
	Gibbs	0.085	1.657	0.791	0.317	2.517	n/a	n/a	n/a	n/a
	Greedy PC	0.0	0.007	0.012	0.028	0.117	1.6947	0.835	4.5333	12.32
Missing	TbM	0	0	0	0	0	0	0	0	0
	Gibbs	0	0	0	2	7	n/a	n/a	n/a	n/a
	Greedy PC	0	0	0	0	1	1	0	24	32
Number of tracks	TbM	12	24	40	65	75	137	124	194	284
	Gibbs	10	20	30	38	43	n/a	n/a	n/a	n/a
	Greedy PC	10	20	30	40	49	50	106	137	179

The measurements are classified into general, long-term, and high-density results. Each experiment was done for N targets and T frames (second row). Occlusions are given in the third row: the absolute numbers sum all occlusions in all views, and relative occlusions indicate the mean number of occlusions per camera and frame. In row max $\lvert\mathcal{C}\rvert$, the size of the biggest cluster in $\mathfrak{S}_{\mathcal{P}}$ (without PC) and $\mathfrak{S}^*_{\mathcal{P}}$ (with PC) is given for Gibbs and Greedy PC, respectively. $\lvert\mathfrak{S}_{\mathcal{P}}\rvert$ indicates only one cluster $\mathcal{C}_{(t-\delta:t),v} = \mathcal{C}_{T,N}$ including all sequences of configurations. Time measurements are given in seconds, besides the entries >4 h highlighting computational time longer than 4 h. Since we focus on biological applicability, we do not discuss tracking results with a computational time >4 h (entries are marked by n/a). In the last row, the E_{ca} measurements are given in combination with the missing flies and the absolute number of calculated tracks.

spondences are unique and latter optimization is only necessary for a few small clusters (compare to max $\lvert\mathcal{C}\rvert$ in Table 1).

This can be observed in all movies besides the movie with 20 flies: the maximum cluster contains 19,024 pairings and 2,140 pairings without and with projection consistency, respectively (Table 1). Thus, the tracking time increases for both GCS approaches. However, the greedy selection is still able to calculate sufficient tracks, whereas Gibbs sampling results in less reliable results. The reason is that Gibbs needs to sample one sequences of configurations in a cluster containing almost 20,000 pairings covering 18 targets for 843 frames. Given one wrong correspondence selection prevents Gibbs from converging in the global optimum. Since we sampled for 10,000 iterations, this was not possible in reasonable time.

This coherence is also observable in the long-term and high-density experiments. In contrast to Zou et al. [23], we merge overlapping clusters for both joint time steps and joint targets (see Section 2.2.3) to guarantee a global search space. Thus, given very dense situations with hundreds of flies, the natural segmentation of the clusters is no longer available. In all measurements, the cluster size

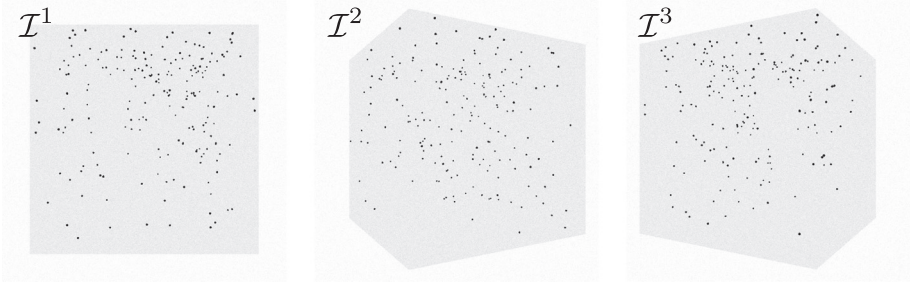

Figure 5 Example of a noisy image triplet generated by the swarm simulator.

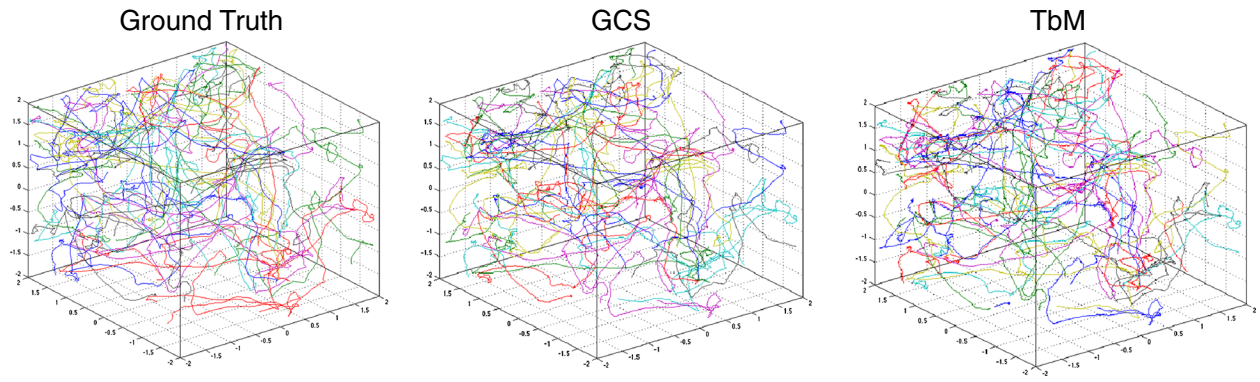

Figure 6 Tracking results of 50 flies over 1,000 time steps and ground truth. The GCS image was made with projection consistency and greedy selection.

of the Gibbs approach was equivalent to the overall search space so that $|\mathcal{C}_{(t-\delta:t),\nu}| = |\mathcal{C}_{T,N}| = |\mathfrak{S}_{\mathcal{P}}|$. The latter optimization must therefore sample one sequence of configurations out of $k^{\nu(\delta+1)}$ ($k \leq \nu$ and $\nu \to N$, $\delta \to T$) possible sequences (compare to Section 2.2.4) statistically. This is why Gibbs sampling requires millions of sampling steps to calculate stable results [37] which was neither shown in [23] nor possible in our data for thousands of frames in reasonable time. Since algorithms requiring more than 4 h for only a few seconds movie length are not suitable for biological applications, we neglect these tracking results in Table 1 (indicated by n/a). Thus, Gibbs sampling for high-density or long-term situations is more interesting from a theoretical point of view [24].

4.2 TbM vs. GCS

Obviously, the Greedy PC approach has the best overall performance int the general experiments. The TbM approach is between the Greedy and Gibbs solution. Optimization of GCS without PC and via Gibbs leads to the worst results with irregular E_{ca} values.

If the number of flies increases, the E_{ca} increases for both TbM and Greedy tracking (compare to Figure 7a). Since both measurements increase proportional to the number of wrong correspondences between views N_c and wrong associations over time N_a, these values are examined in Figure 7b.

As apparent, an increasing N_c value leads to high error measurements for both the TbM and the Greedy PC

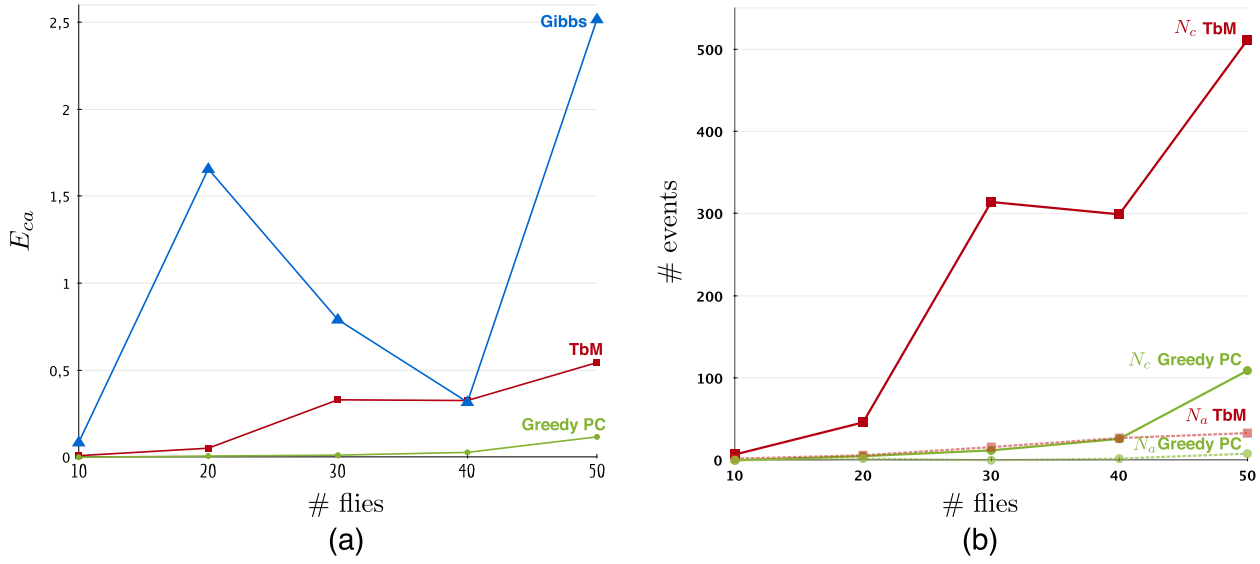

Figure 7 Resultant E_{ca} and N_c and N_a values. (a) Resultant E_{ca} value plotted for all cases from Table 1. **(b)** Resultant N_c and N_a values plotted for the TbM and the Greedy PC tests.

approach. The main reason for wrong or missing stereo correspondences is caused by occlusions. The more flies are located in the chamber, the more occlusions arise during blob detection (compare to Table 1). Especially in latter frames, occlusions arise very frequently because of the negative geotaxis (compare to Section 3.1). During the movie with 50 flies, up to 4 flies are occluded for several frames in camera 1, for example. Thus, even in situations with up to 50 flies, the target density is comparatively high. Since TbM and Greedy PC try to overcome this events using the projection consistency, both have much lower N_c than the two camera tracking solutions. Gibbs has up to 1,677 wrong stereo correspondences (data not shown). Therefore, it is not able to calculate the global optimum even after 10,000 sampling steps.

In other words, the overall search space $\mathfrak{S}^*_{\mathcal{P}}$, containing more than 1,000 occlusions, cannot be sampled sufficiently because of the growth of the clusters (Table 1). However, Greedy PC benefits from the previously calculated overall search space $\mathfrak{S}^*_{\mathcal{P}}$: since all possible pairings and sequences of configurations are used for coast calculations, ambiguities caused by occlusions can be corrected more frequently.

4.3 Long-term tracking

In the long-term experiment, 50 flies were tracked for 3,000 frames. Gibbs failed in this experiments because the size of the clusters $\mathcal{C}$ increases drastically for 3,000 frames. Thus, only TbM and Greedy PC were able to track during this experiment.

In long-term movies, the TbM approach can achieve better results than the Greedy PC algorithm (Table 1). The reason for this inversion compared to the 1,000 frame results is that the size of the clusters $|\mathcal{C}|$ increases to much for 3,000 frames. Thus, the probability of getting a local optimum via greedy selection increases accordingly. However, tracking accuracy is still convenient in the Greedy PC approach. On the other hand, TbM, as a causal method, is not affected by the length of tracking sequences.

In contrast to the TbM approach, GCS can miss targets during optimization (see Table 1). However, projection consistency reduces the amount of missed targets. Furthermore, GCS optimization leads to less fragmented trajectories than the TbM approach. Whereas TbM results in 137 trajectories ($N_a = 99$) for 50 flies (over 3,000 frames), the Greedy PC approach calculates 49 complete tracks of 50 tracks in total ($N_a = 7$). If complete trajectories are required (i.e., identity of the flies must remain over time), Greedy PC is recommended but with the possibility of loosing flies.

4.4 High-density tracking

To evaluate the behavior of both tracking approaches, we tracked up to 200 flies. Similar to the long-term experiment in Section 4.3, we limit our comparisons to Greedy PC and TbM.

Table 1 highlights the measurements for 100, 150, and 200 targets. We decreased the number of frames for the movies with 150 and 200 flies to limit the size of $\mathfrak{S}^*_{\mathcal{P}}$. For up to 100 targets and 200 frames, Greedy correspondence selection is more accurate than TbM. However, given more than 100 targets, resulting in very high fly densities, the TbM outperforms the GCS approach. Most importantly, there are no missing targets in the probabilistic approach, whereas Greedy was not able to find trajectories for all flies. Furthermore, TbM can achieve better overall accuracy in high-density situations in less tracking time. The only drawback of the probabilistic tracking is again the fragmentation of the trajectories: TbM calculates more tracks than Greedy PC resulting in many identity changes.

5 Conclusion

In this paper, we discussed two tracking paradigms for identical appearing objects such as *Drosophila melanogaster* in 3D. One paradigm is based on a probabilistic approach conducting tracking and matching alternatively [35]. The other paradigm constructs a global search space over all targets and time steps, which is optimized in a second step [23].

Due to the high complexity of the second GCS paradigm, we introduced two improvements, namely projection consistency and greedy optimization. Especially, the projection consistency is able to reduce the overall complexity and thus improve the tracking results without yielding into local optima. Since Gibbs sampling, used for GCS optimization in [23], needs thousands of iterations to guarantee good results, our greedy selection scheme outperforms Gibbs sampling. However, a global result cannot be guaranteed via greedy optimization.

We demonstrated several advantages and disadvantages of both the TbM and GCS approach. Thus, the decision which approach to use must be done carefully. If the identity of the flies is not important, TbM can be used to track for several thousands of frames (compare to Section 4.3). All flies were detected, but the trajectories were fragmented due to occlusions.

The GCS approach was not able to track all flies in all experiments: only 49 of 50 flies were detected. On the other hand, the trajectories of the detected flies were less fragmented (compare to Section 4.2). In addition, the GCS was able to solve collisions and occlusions more frequently because of the global search space. This leads to the higher tracking accuracy illustrated in Figure 7a. If dozens of flies must be tracked for a comparatively short period, GCS outperforms TbM tracking. For very long sequences, it is the other way around.

If high fly densities are needed for a comparatively long period, the size of the global search space prevents Gibbs optimization, because it requires too many sampling iterations. In addition, greedy tracking quality decreases drastically compared to TbM (see Section 4.4). Thus, without further reductions of the global search space, probabilistic tracking is the preferable paradigm in high-density experiments.

TbM could be optimized in terms of tracking accuracy, whereas GCS could be optimized for longer tracking durations and higher target densities. Possible improvements for the TbM approach are discussed in [35]. Here, we want to focus on improvements of the GCS approach.

Currently, we use the third camera only to correct mismatches between cameras 1 and 2. The optimization scheme is still executed on pairings. Since pairwise comparison in a triplet would further reduce the search space, all optimization steps could be done on three image points.

The kinetic model given by Equation 8 is also a current drawback of the GCS approach. Only motion form $(t-1)$ is considered for time step t. Thus, a more appropriate motion model would further improve the accuracy of GCS tracking.

Currently, we are developing a three-camera real-world setup to capture movies of adult *Drosophila* flies. Thus, we are going to test both algorithms on real video sequences, comparable to the synthetic data introduced above.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors thank *S. Strothoff* for the discussion and help throughout the project. Furthermore, we would like to thank the anonymous reviewers for the helpful annotations and suggestions for improvements. We acknowledge the support by the Deutsche Forschungsgemeinschaft and Open Access Publication Fund of University of Münster.

Author details

[1]Department of Mathematics and Computer Science, University of Münster, Einsteinstraße 62, Münster 48149, Germany. [2]Department of Computer Science, University of Auckland, Tamaki Innovation Campus, Glen Innes, Morrin Road, Auckland 1142, New Zealand. [3]Department of Neuro and Behavioral Biology, University of Münster, Badestraße 9, 48149 Münster, Germany.

Received: 31 January 2013 Accepted: 19 September 2013
Published: 31 October 2013

References

1. DM Gohl, MA Silies, XJ Gao, S Bhalerao, FJ Luongo, CC Lin, CJ Potter, TR Clandinin, A versatile in vivo system for directed dissection of gene expression patterns. Nat. Methods. **8**(3), 231–237 (2011)
2. A Katsov, Motion processing streams in Drosophila are behaviorally specialized. Neuron. **59**(2), 322–335 (2008)
3. B Risse, S Thomas, N Otto, T Löpmeier, D Valkov, X Jiang, C Klämbt, FIM a novel FTIR-based imaging method for high throughput locomotion analysis. PloS one. **8**(1), e53963 (2013)
4. MB Sokolowski, *Drosophila*: genetics meets behaviour. Nat. Rev. Genet. **2**(11), 879–890 (2001)
5. A Gomez-Marin, N Partoune, GJ Stephens, M Louis, Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. PloS one. **7**(8), e41642 (2012)
6. D Grover, J Yang, D Ford, S Tavaré, J Tower, Simultaneous tracking of movement and gene expression in multiple *Drosophila melanogaster* flies using GFP and DsRED fluorescent reporter transgenes. BMC Res. Notes. **2**(1), 58 (2009)
7. S Lahiri, K Shen, M Klein, A Tang, E Kane, M Gershow, P Garrity, ADT Samuel, Two alternating motor programs drive navigation in *Drosophila* larva. PloS one. **e23**(8), 180 (2011)
8. J Tao, R Klette, Tracking of 2D or 3D irregular movement by a family of unscented Kalman filters. J. Inf. Convergence Commun. Eng. **10**(3), 307–314 (2012)
9. H Dankert, L Wang, ED Hoopfer, DJ Anderson, P Perona, Automated monitoring and analysis of social behavior in Drosophila. Nat. Methods. **6**(4), 297–303 (2009)
10. J Martin, A portrait of locomotor behaviour in Drosophila determined by a video-tracking paradigm. Behav. Process. **67**(2), 207–219 (2004)
11. TA Ofstad, CS Zuker, MB Reiser, Visual place learning in Drosophila melanogaster. Nature. **474**(7350), 204–207 (2011)
12. J Colomb, L Reiter, J Blaszkiewicz, J Wessnitzer, Open source tracking and analysis of adult Drosophila locomotion in Buridan's paradigm with and without visual targets. PloS one. **7**(11), e42247 (2012)
13. D Valente, I Golani, PP Mitra, Analysis of the trajectory of Drosophila melanogaster in a circular open field arena. PloS one. **2**(10), e1083 (2007)
14. SN Fry, M Bichsel, P Müller, D Robert, Tracking of flying insects using pan-tilt cameras. J. Neurosci. Methods. **101**(1), 59–67 (2000)
15. A Poore, Multidimensional assignment formulation of data association problems arising from multitarget and multisensor tracking. Comput. Optimization Appl. **3**, 27–57 (1994)
16. RE Burkard, M Dell'Amico, S Martello, *Assignment Problems*, 1st edn (Society for Industrial Mathematics, Philadelphia, 2009)
17. LF Tammero, MH Dickinson, The influence of visual landscape on the free flight behavior of the fruit fly Drosophila melanogaster. J. Exp. Biol. **205**(Pt 3), 327–343 (2002)
18. H Du, D Zou, YQ Chen, Relative epipolar motion of tracked features for correspondence in binocular stereo, in *ICCV Conference Proceedings,* Rio de Janeiro, October 2007 (IEEE, Piscataway, 2007), pp. 1–8
19. D Engelmann, C Garbe, M Stohr, P Geißler, F Hering, B Jahne, Stereo particle tracking, in *8th International Symposium on Flow Visualisation,* Sorrento, 1–4 September (1998), pp. 1–10
20. HS Wu, Q Zhao, D Zou, YQ Chen, Acquiring 3D motion trajectories of large numbers of swarming animals, in *IEEE 12th International Conference on Computer Vision Workshops,* Kyoto, September to October 2009 (IEEE, Piscataway, 2009), pp. 593–600
21. D Grover, J Tower, S Tavaré, O fly, where art thou. J. R. Soc. Interface. **5**(27), 1181–1191 (2008)
22. AD Straw, K Branson, TR Neumann, MH Dickinson, Multi-camera real-time three-dimensional tracking of multiple flying animals. J. R. Soc. Interface. **8**(56), 395–409 (2011)
23. D Zou, Q Zhao, HS Wu, YQ Chen, Reconstructing 3D motion trajectories of particle swarms by global correspondence selection, in *IEEE 12th International Conference on Computer Vision,* Kyoto, September to October 2009 (IEEE, Piscataway, 2009), pp. 1578–1585
24. R Ardekani, A Biyani, JE Dalton, JB Saltz, MN Arbeitman, J Tower, S Nuzhdin, S Tavaré, Three-dimensional tracking and behaviour monitoring of multiple fruit flies. J. R. Soc. Interface. **10**(78), 20120–547 (2013)
25. F Pereira, H Stüer, EC Graff, M Gharib, Two-frame 3D particle tracking. Meas. Sci. Technol. **17**(7), 1680–1692 (2006)
26. D Theriault, Z Wu, N Hristov, S Swartz, K Breuer, T Kunz, M Betke, Reconstruction and analysis of 3D trajectories of Brazilian free-tailed bats in flight (2010)
27. Z Wu, NI Hristov, TL Hedrick, TH Kunz, M Betke, Tracking a large number of objects from multiple views, in *2009 IEEE 12th International Conference on Computer Vision,* Kyoto, September to October 2009 (IEEE, Piscataway, 2009), pp. 1546–1553
28. A Veeraraghavan, M Srinivasan, R Chellappa, E Baird, R Lamont, Motion based correspondence for 3D tracking of multiple Dim objects. IEEE Int. Conf. Acoustics Speech Signal Process (ICASSP). **2**, pp. II (2006)

29. NT Ouellette, H Xu, E Bodenschatz, A quantitative study of three-dimensional Lagrangian particle tracking algorithms. Exp. Fluids **40**(2), 301–313 (2005)
30. KJ Kohlhoff, TR Jahn, DA Lomas, CM Dobson, DC Crowther, M Vendruscolo, The iFly tracking system for an automated locomotor and behavioural analysis of *Drosophila melanogaster*. Integr. Biol. (Camb). **3**(7), 755–760 (2011)
31. S Zou, P Liedo, L Altamirano-Robles, J Cruz-Enriquez, A Morice, DK Ingram, K Kaub, N Papadopoulos, JR Carey, Recording lifetime behavior and movement in an invertebrate model. PloS one. **6**(4), e18151 (2011)
32. S Butail, N Manoukis, M Diallo, JM Ribeiro, T Lehmann, DA Paley, Reconstructing the flight kinematics of swarming and mating in wild mosquitoes. J. R. Soc. Interface. **9**(75), 2624–2638 (2012)
33. DH Kelley, NT Ouellette, Emergent dynamics of laboratory insect swarms. Sci. Rep. **3** (2013)
34. A Attanasi, A Cavagna, L Del Castello, I Giardina, A Jelic, S Melillo, L Parisi, E Shen, E Silvestri, M Viale, Tracking in three dimensions via multi-path branching. CoRR abs/1305.1495 (2013)
35. J Tao, B Risse, X Jiang, R Klette, 3D Trajectory estimation of simulated fruit flies, in *Proc 27th IVCNZ*, (Dunedin, 26–28 November 2012)
36. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University Press, Cambridge, 2004)
37. S Geman, D Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**(6), 721–741 (1984)
38. EA Wan, R van der Menve, The unscented Kalman filter for nonlinear estimation, in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000,* Lake Louise, Alta, October 2000 (IEEE, Piscataway, 2000), pp. 153–158
39. JH Marden, MR Wolf, KE Weber, Aerial performance of Drosophila melanogaster from populations selected for upwind flight ability. J. Exp. Biol. **200**(Pt 21), 2747–2755 (1997)
40. JW Gargano, I Martin, P Bhandari, MS Grotewiel, Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila. Exp. Gerontol. **40**(5), 386–395 (2005)

doi:10.1186/1687-5281-2013-57
Cite this article as: Risse *et al.*: Comparison of two 3D tracking paradigms for freely flying insects. *EURASIP Journal on Image and Video Processing* 2013 **2013**:57.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com