Response to the review comments of reviewer 4

Comments and Suggestions for Authors

Comment 1: The manuscript's language should be written in passive voice as it is common in the scientific literature. For example, “it is maintained” or “finite sets are mixed” rather than “we maintain” or “we mix”.

Response:

Dear Reviewer,

Thank you for your review and suggestions. We appreciate your guidance, especially the recommendation to use the passive voice, which is a common practice in scientific writing. We will make the necessary revisions to the paper to adhere to this writing convention.

In the revisions, we will employ the passive voice, such as changing "we maintain" to "it is maintained" and "we mix" to "finite sets are mixed," to ensure the paper's language style aligns with scientific literature. We will also review other parts of the paper and address similar issues.

We are grateful for your guidance, which will contribute to the quality and professionalism of our paper. If you have any further suggestions or require additional revisions, please do not hesitate to let us know, and we will do our best to accommodate your requests.

Comment 2: In the Conclusion part, conclusions of the work should be strongly stated rather than a general overview of the study.

Response:

Dear Reviewer,

Thank you very much for your valuable suggestions. We understand your advice regarding the conclusion section, specifically the recommendation to state the conclusions of the work more emphatically rather than providing a general overview of the research. Based on your feedback, we will make the following revisions:

> Our research intends to integrate the advantages of detection-based visual multi-object tracking algorithms with RFS-based visual multi-object tracking algorithms in order to improve the performance of drone aerial video multi-object tracking. In addition, a more comprehensive, robust, and efficient integrated multi-object tracking algorithm is proposed by modelling object motion information. Our method has produced a number of novel results in the field of drone aerial multi-object tracking, DB-Tracker achieved MOTA of 37.4% and 46.2% on the VisDrone and UAVDT datasets respectively, and it offers a practical solution to the multi-object tracking problem in real-world scenarios. Simultaneously, our method has improved significantly in a variety of datasets and tough settings. Our study contributes new ideas and methodologies to the advancement of the

field of unmanned aerial vehicle visual tracking. In the future, we will persistently pursue enhancements and optimizations of the algorithm, aiming to elevate the performance of multi-object tracking even further. Moreover, we are committed to extending the applicability of our research findings across a broader spectrum of application domains. Moving forward, our dedication to improving and fine-tuning the algorithm remains unwavering, with the goal of achieving superior performance in multi-object tracking tasks. Additionally, we are eager to leverage the insights gained from this study to benefit a wider array of application areas.

Through these modifications, we will make the conclusion section more prominent and targeted, in order to better convey the importance and findings of the research. Thank you once again for your suggestions, and we will continue to strive to improve our paper.

If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to meet your requirements.

Comment 3: In the manuscript, consecutive references should be separated with a comma (i.e. [1, 2]). For example, in line 34 or in line 53.

Response:

Dear reviewer,

Thank you for pointing out the citation format issue in the paper. We greatly appreciate your attention and suggestions. Based on your advice, we have made the following modifications:

1. In the paper, we will ensure to use commas to separate consecutive references in the citation, in order to comply with the correct citation format. For example, we will correct "[1 2]" to "[1, 2]".

2. We have carefully reviewed the entire paper to ensure consistency in the citation format and made necessary corrections where needed.

The modified sections are as follows:

Drones have been rapidly promoted as electronic communication technology has advanced due to their advantages of mobility, ease of operation, and low cost, compensating for the loss of information caused by traditional means due to weather and time constraints. At the same time, the high agility of drones, when compared to fixed cameras, can make the scope of aerial video more flexible and diverse. Drone video data is particularly informative in terms of content and time, promoting the growing importance of drone aerial video in numerous sectors of object detection and tracking [1,2]. However, as compared to multi-object recognition and tracking jobs in ordinary view videos, drone aerial video faces several obstacles, such as frame degradation, uneven object distribution density, tiny object size, and real-time performance, which have piqued the interest of academics in recent years. There has been a lot of interest and research in the world and industry [3].

Many academics have developed various multi-object tracking (MOT) algorithms for drone aerial video in recent years, with the most frequent way being multi-object tracking based on visual object identification, which has become the most popular multi-object tracking task for drone aerial video valid frame. The detection-based tracking process is often divided into two parts: (1) a motion model and state estimation to anticipate the trajectory's bounding box in subsequent frames; and (2) linking new frame detections with

> the present trajectory [4,5]. For dealing with association tasks, there are two main approaches: (1) Object appearance models and re-identification problems; (2) Object localization, specifically the intersection-over-union ratio between predicted trajectory bounding boxes and detection bounding boxes [6]. Both approaches use distances to quantify associations and treat the association issue as a global assignment problem. However, because to the uncertainty of object recognition and the mutual occlusion and overlap of objects, these approaches may experience certain issues when dealing with complicated scenarios [7,8].

These modifications will help improve the accuracy and professionalism of the paper. Thank you again for your suggestions, and please let us know if you have any further suggestions or require additional modifications. We will do our best to meet your requirements.

Comments on the Quality of English Language

Comment 1: Moderate editing of English language required.

Thank you for your review comments. We have made moderate English edits to ensure more accurate and fluent language expression in the article. We have carefully reviewed and corrected grammar and spelling errors, as well as unclear sentences. If you have any questions or need further guidance, please feel free to let us know. Thank you again for your valuable feedback.

Thank you for your attention.

Best regards,

Yubin Yuan