2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Photometric Consistency for Precise Drone Rephotography

Hsaun-Jui Chang[1], Tzu-Chun Huang[2], Hao-Liang Xu[1] and Kuan-Wen Chen[1*]

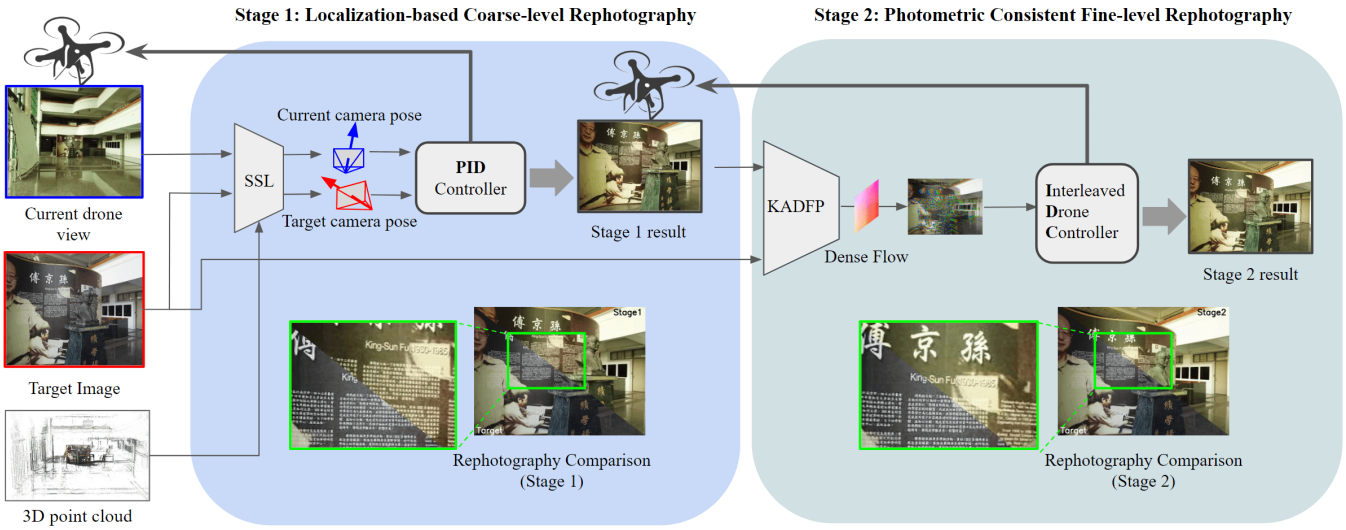


Fig. 1: The operation of the proposed drone rephotography system comprises two stages. Stage 1 involves **Localization-based Coarse-level Rephotography** for initial positioning, and Stage 2 involves **Photometrically Consistent Fine-level Rephotography**. In this stage, the *KADFP* model and an *interleaved drone controller* are used for fine-tuned rephotography. **Rephotography Comparison** showing the combination that the upper right halves of both stages with the lower left of the target image, clearly demonstrate the improvement achieved in Stage 2.

***Abstract*— This paper proposes a precise drone rephotography system for fixed-domain patrolling scenarios. The proposed system integrates computer-vision-based localization and fine-tuned pixel-level dense flow prediction to achieve consistent and precise rephotography images with viewpoints that closely align with those of target images. The proposed Keypoints Alignment Through Dense Flow Prediction (KADFP) model effectively handles challenges arising from lighting variations and background differences. Moreover, a novel flight procedure is implemented in the proposed system. This procedure involves using an Interleaved Drone Controller to alternate between translation and rotation adjustments to ensure smooth flight dynamics during rephotography. Experiments indicated that the proposed system provided considerably more precise rephotography results (error of 4.72 pixels indoors) than did an existing localization approach (error of 35.56 pixels).**

I. INTRODUCTION

Rapid advancements in drone technology have resulted in the gradual diversification of tasks that can be executed by drones and continual improvement of the operational precision of drones. Among the numerous applications of drone technology, precise rephotography is a relatively underdeveloped field. The aim of precise rephotography is to control drones during flight for automatically searching and identifying certain known areas or objects and then capturing identical perspectives to those of target images with high precision. In certain unmanned aerial vehicle (UAV) inspection tasks, it may be required for the UAV to return an image *I'* for comparison with an image *I* captured several days or even years ago. When executing this task, if there is a significant deviation in the angle between the re-captured image *I'* and the previously captured image *I*, subsequent comparison algorithms may need to be designed with complexity. However, if the perspective of image *I'* closely resembles that of image *I*, subsequent comparison algorithms may be designed more straightforwardly, potentially leading to improved accuracy.

Precision rephotography technology offers practical solutions for automated inspections, particularly in hazardous settings like electrical towers and factories. Drones have gained popularity in this field due to their safety, reliability, and cost-effectiveness. This paper introduces a drone rephotography system designed to align the structure of images with previously captured target images. This not only boosts the precision of automated inspections but also reduces manual postprocessing costs. High-similarity image

[1]Hsaun-Jui Chang , Hao-Liang Xu and Kuan-Wen Chen are with the Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan. (E-mail of corresponding author* Kuan-Wen Chen: kuanwen@cs.nycu.edu.tw)
[2]Tzu-Chun Huang is member of Internet of Things Laboratory, Chunghwa Telecom Laboratories Taoyuan, Taiwan (tzuchun@cht.com.tw)

pairs improve accuracy in various computer vision tasks, such as object recognition [1], facial recognition [2], pose estimation [3], depth estimation [4], and image alignment [5]. While commercial drones, like those from DJI [6], set flight routes using precaptured content and GPS data, interference can compromise GPS accuracy in certain environments [7]. As a result, this study introduces a rephotography system that relies solely on visual localization cues.

As mentioned, this paper proposes a drone rephotography system (Fig. 1) based on visual localization for enabling a drone to fly autonomously to the exact location of an original photograph for rephotography. This system allows a drone to perform rephotography by utilizing only its monocular camera and without relying on GPS information or any additional sensors. The operation of the proposed system involves two stages: (1) **Localization-based Coarse-level Rephotography** (Stage 1) and (2) **Photometric Consistent Fine-level Rephotography** (Stage 2). In Stage 1, the Single-Shot Localization (SSL) method [8] is used. Initially, the target image is positioned, after which an approximate localization result is generated by maintaining continuous proximity to the target location. However, even when the drone is considerably close to the target position, it cannot guarantee that the rephotography result would align with the target image, as illustrated in Fig. 1. Consequently, photometric consistent fine-level rephotography, which is a fine-tuning method based on dense flow prediction, is performed. In this process, a model developed in the present study, namely the Keypoints Alignment from Dense Flow Prediction (KADFP) model, is used to execute pixel-level dense flow prediction between the target image and the current drone view. Subsequently, a keypoints detector, (used SIFT [9]), is used to select the key flow. An interleaved drone controller is then used to iteratively fine-tune the drone's translation and rotation to ensure that the newly captured image is aligned with the target image. This step ensures that the desired target image and rephotographed image have photometric consistency.

In summary, the contributions of this study are outlined as follows:

- A drone rephotography system is proposed for patrolling within a fixed domain. This system integrates localization based on computer vision and pixel-level fine-tuning based on dense flow prediction. The proposed system achieves highly precise rephotography results (error: 4.72 pixels), and its precision is considerably greater than that of an existing visual localization method (35.56 pixels).
- The proposed KADFP model can accurately predict the dense flow correspondences for image pairs with lighting variations or partial background differences. Moreover, we used the keypoints alignment error to determine the proximity between image pairs.
- A flight procedure that involves alternating between translation and rotation by using an interleaved drone controller is proposed. This procedure enhances the flight smoothness during rephotography, which results in the rephotography results resembling the target images more closely.

II. RELATED WORK

In today's evolving drone technology landscape, most commercial drone localization solutions primarily depend on external navigational aids. These aids, which predominantly comprise systems such as the GPS [15] [16], ultra-wideband technology, and inertial measurement units, enable drones to achieve stable and reliable operation in open spaces. However, limitations exist in drone navigation based on these aids, especially in areas with obstructions or with signal interference.

Vision-based methods offer enhanced flexibility and accuracy in drone navigation compared to other aids. Consequently, their use in drone autopilot system development has surged in both academia and industry. For instance, DroneTalk [34] discusses a drone's visual self-localization for efficient urban package delivery. Additionally, visual homing [32] [33], another innovation, enables drones with cameras to use computer vision for recognizing landmarks and autonomously returning to their starting points.

In visual homing tasks, the drone predominantly uses images captured during its takeoff for navigation. By contrast, in rephotography, the drone might need to align its position by using an image that was captured under entirely different lighting or weather conditions from the current conditions. For example, taking an image of the target during a clear day, and later achieving accurate rephotography even in cloudy weather or at night. Such variations between the conditions corresponding to the target image and the current conditions can cause considerable challenges for existing navigation algorithms. To the best of our knowledge, no study has addressed this challenge. Consequently, developing innovative solutions for the aforementioned problem is imperative because with the evolution of drone technology, the need for specialized visual tasks, such as precise drone rephotography, is expected to increase considerably.

III. METHODS

The overall architecture of the proposed system is depicted in Fig. 1. In Stage 1 (Localization-based Coarse-level Rephotography), a point-cloud model of the inspection area is constructed, and the SSL method, which is a hierarchical localization method proposed by [8], is used for six-degree-of-freedom (DoF) camera positioning. On the basis of the localization results for the target image, which are obtained in advance, a proportional–integral–derivative (PID) controller [26] moves the drone to an approximate rephotography position. In Stage 2 (Photometric Consistent Fine-level Rephotography), this position is refined using the dense flow predicted by the KADFP model. This model predicts the pixel-level dense flow between the target image and the current drone view. The key flow is extracted from the dense flow by using a keypoints detector (using SIFT [9]),

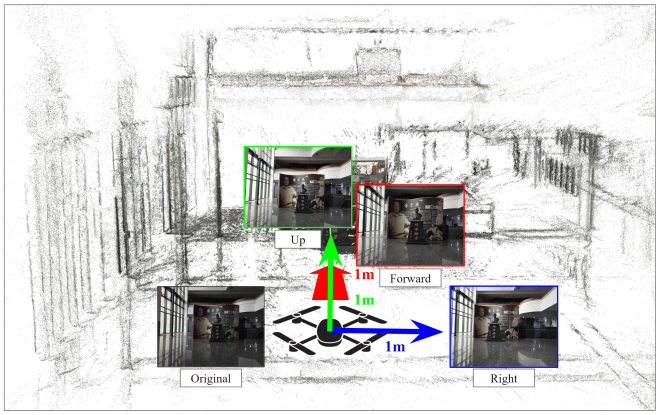


Fig. 2: These images comprise a original image and three others, each taken after moving one meter in one of the three directions along the drone's flight axes.

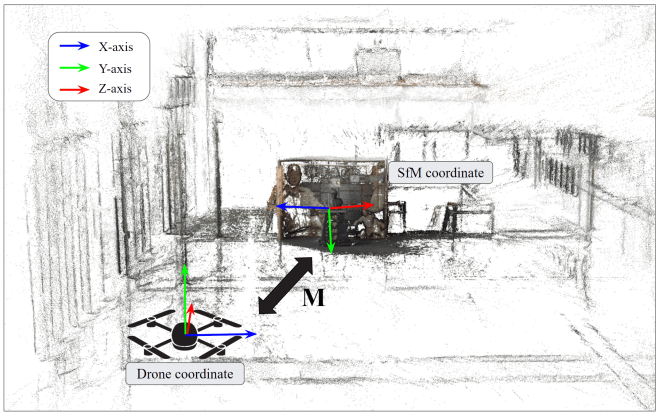


Fig. 3: M is used for Coordinate transformation, we can see the difference between SfM coordinate system and drone's.

and fed into the Interleaved Drone Controller for flight fine-tuning. This controller switches the drone movement between translation and rotation for achieving precise rephotography.

A. Stage 1: Localization-based Coarse-level Rephotography

Initially, the structure-from-motion method [27] is employed to create a three-dimensional (3D) model to generate a point cloud and establish a 3D coordinate system. Additionally, to enhance feature matching, deep-neural-network-based descriptors are adopted [30]. For efficient image localization, the hierarchical localization approach [8] is employed. In this approach, a deep neural network is used for image retrieval to narrow the search scope. Features are then matched between the query and retrieved images for establishing 2D–3D correspondences. These correspondences are used to apply PnP [28] to acquire the camera pose of the query image, and RANSAC [29] is utilized for outlier removal.

After the camera poses of the current drone view and target image are acquired, their relative poses are calculated, and a PID controller [26] is used to generate flight control commands. Since 3D point cloud models lack scale information, performing localization tasks directly can result in errors. To address this, we derive a transformation matrix to convert the camera pose into the drone's coordinate system, ensuring accurate flight control. During data collection, we include four additional images representing the '*original*', '*right*', '*up*' and '*forward*' directions, as seen in Fig.2. These images consist of an original image and three others, each taken after moving one meter in one of the three directions along the drone's flight axes. The purpose is to align the camera pose results obtained through Single-Shot Localization with the coordinate system oriented to the drone. This alignment enables us to use the difference between the current drone's camera pose and the target image pose for drone control. We employ linear algebra coordinate transformation for this specific control method, as shown in the following formula:

$$M \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + X_{\mathrm{o}} = \begin{bmatrix} X_{\mathrm{right}} & X_{\mathrm{up}} & X_{\mathrm{forward}} \end{bmatrix} \quad (1)$$

Here, M represents the rotation matrix between these two coordinate systems. X_{o} represents the pose results of the original image in the coordinate system based on Structure from Motion (SfM). X_{right}, X_{up}, and X_{forward} represent the pose results based on SfM coordinate system obtained after moving one meter in the right, up, and forward directions along the drone's flight axis. First, we calculate the transformation matrix M, and then we perform coordinate transformations using the formula above.

$$X_{\mathrm{new}} = \mathrm{inv}(M) \cdot (X_{\mathrm{old}} - X_o) \quad (2)$$

X_{old} represents the actual pose results based on the SfM coordinate system obtained during the actual flight, and X_{new} is the pose results based on the drone's coordinate system after the coordinate transformation. Refer to Fig. 3 for a visualization of this coordinate transformation.

B. Stage 2: Photometric Consistent Fine-level Rephotography

Because of the inherent limitations related to algorithmic precision and the processing of large scenes, Stage 1 might result in the production of imprecise camera poses. To refine the rephotography results obtained in Stage 1, Stage 2 is executed. In Stage 2, the predicted dense flow between the current drone view and the target image is used to obtain pixel-level correspondences. Subsequently, the key flow is selected from the dense flow for performing fine-level drone position adjustments.

1) KADFP Model: Fig. 4 illustrates the architecture of the proposed KADFP model. In contrast to traditional optical flow models, the KADFP model considers lighting and scene variations between the current drone view and the target image. This model is based on the RAFT [11] architecture and incorporates elements from the LIFE [12] method.

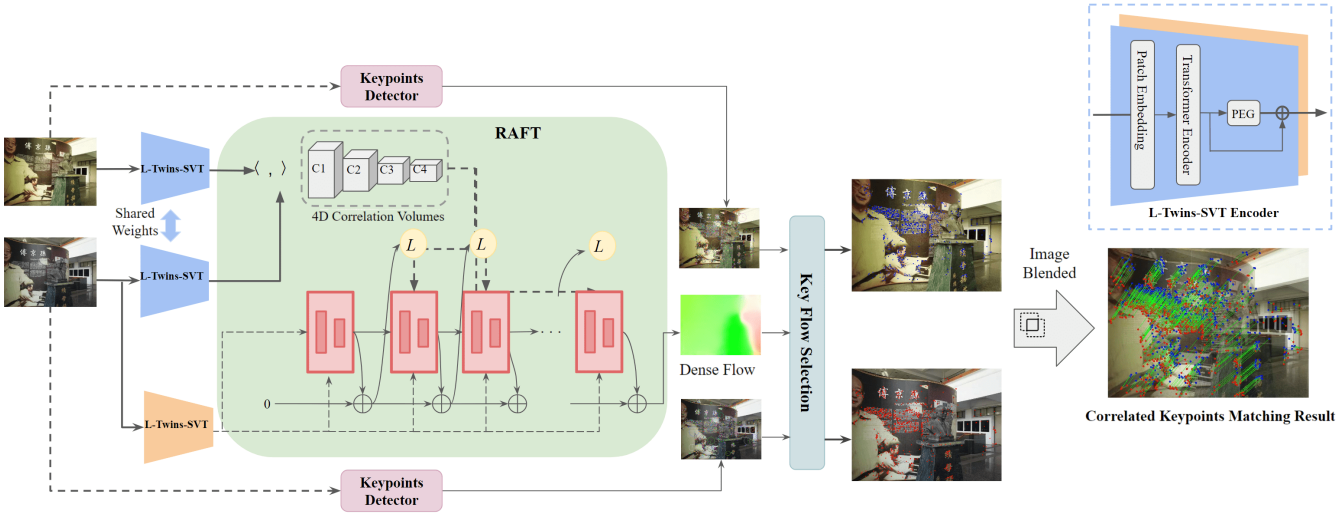


Fig. 4: Architecture of the KADFP model. This dense flow prediction model is based on the RAFT architecture and includes an image encoder based on the Transformer architecture. The **Correlated Keypoints Matching Result** is the result of blending two images is presented. The blue points represent keypoints from the current drone view, the red points represent keypoints from the target image, and the green lines indicate the key flow.

In the KADFP model, the convolutional neural network (CNN)-based image encoder of the RAFT architecture is replaced with an L-Twins-SVT structure (top-right panel in Fig. 4), which is derived from the Twins-SVT [13] framework. The L-Twins-SVT structure improves image information capture while maintaining low computational costs. Moreover, the *Symmetric Epipolar Distance (SED) loss* and *the synthetic dense flow regularization loss* proposed by LIFE [12] are used to enable the execution of weakly supervised training using only camera poses from the MegaDepth [31] dataset. To improve the precision of flow prediction in keypoints regions and thus enhance the model performance, a **keypoints loss** based on the keypoints detector (SIFT [9]) is used to calculate the keypoints prediction error.

Although the developed prediction model generates dense flow results, the direct utilization of the same flows for all points might result in scene disparities, particularly in background and occluded areas. To address this problem, a key flow selection module was developed in this study. After the initial prediction of dense flows, this module utilizes the SIFT [9] to identify keypoints from the target image and current drone view. If the start and end points of a given dense flow are detected as keypoints, the flow is deemed to be trustworthy and designated as "key flow."

For each pair of corresponding points in the key flow, the average distance error, which is termed the **keypoints alignment error** $E^{\text{cur}}_{\text{kpa}}$, is calculated in pixels. This metric can be used for model performance assessment and for result interpretation in subsequent evaluations. The term $E^{\text{cur}}_{\text{kpa}}$ is calculated using the following formula:

$$E^{\text{cur}}_{\text{kpa}} = \frac{1}{|K|} \sum_{(X_{\text{drone}}, X_{\text{target}}) \in K} d(X_{\text{drone}}, X_{\text{target}}) \qquad (3)$$

where K denotes all the point pairs within the key flow;

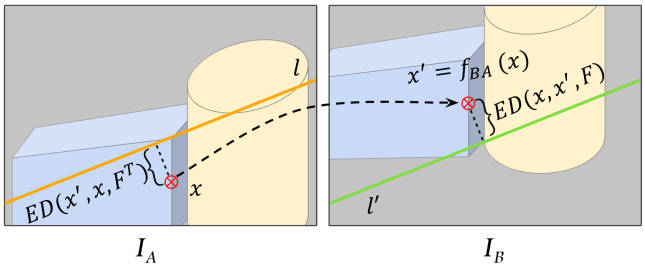


Fig. 5: Symmetric Epipolar Distance.

X_{drone} and X_{target} denote the 2D coordinates (x, y) of the selected keypoints in the image, respectively; and $d(,)$ denotes the Euclidean distance between X_{drone} and X_{target}.

The term **Total Loss** represents a combination of three losses: the SED loss, synthetic dense flow regularization loss [12], and keypoints loss. This term is expressed as follows:

$$L_{Total} = L_{\text{SED}} + L_{BiT} + \alpha * L_{KP} \qquad (4)$$

To increase the influence of the keypoints loss in the model and enhance the accuracy of keypoints prediction, this loss is assigned a weight, which is denoted as α and was set to 2 in this study.

Symmetric Epipolar Distance (SED) loss [12] is a technique based on epipolar geometry used to assess the accuracy of dense flow predictions from one image (I_A to another I_B). It computes a fundamental matrix (F) that relates I_A and I_B based on their camera parameters and relative pose. In essence shown in Fig. 5, it ensures that a pixel (x) in I_A corresponds to a specific point on an epipolar line (l') in I_B, and vice versa.The SED loss aims to measure how well the predicted flow from IA to I_B ($x' = f_{B \leftarrow A}(x)$) aligns with the epipolar line (l') in I_B.

Likewise, it evaluates how well the reverse flow aligns with the epipolar line derived from x' in I_A. The SED is the sum of these two epipolar distances:

$$\text{SED}(x, x', F) = \text{ED}(x, x', F) + \text{ED}(x', x, F^T) \quad (5)$$

where F is the fundamental matrix that relates images I_A and I_B, ED is the epipolar distance for a particular point, and SED is the symmetric epipolar distance. The cumulative SED loss over all pixels in image I_A is expressed as follows:

$$L_{\text{SED}} = \sum_{x_i \in S} \text{SED}(x_i, f_{B \leftarrow A}(x_i), F) \quad (6)$$

Synthetic dense flow regularization is a method designed to enhance flow prediction accuracy. In this method, the SED loss is increased by performing synthetic transformations to improve pixel-to-pixel correspondence accuracy. Random affine or thin-plate spline transformations (T) are generated for each image pair to transform image I_B into a synthetic image I'_B; thus, a paired set $< I_B, I'_B >$ with precise pixel correspondences is obtained. The bidirectional geometric transformation (BiT) loss is expressed as follows:

$$\begin{aligned} L_{BiT} = & \sum_{x_i \in S_B} \| f_{B' \leftarrow B}(x_i) - T(x_i) \|_1 \\ & + \sum_{x_i \in S_{B'}} \| f_{B \leftarrow B'}(x_i) - T^{-1}(x_i) \|_1 \end{aligned} \quad (7)$$

where the transformation T and the corresponding inverse transformation T^{-1} delineate precise pixel correspondences between the synthetic image pairs. The terms S_B and $S_{B'}$ represent valid pixel locations in images I_B and I'_B, respectively. Any flow predictions outside the target image's observable area are deemed invalid. For a comprehensive explanation and further details, readers can see the original research paper.

The **keypoints loss** is a loss function for enhancing the accuracy of predicting keypoints in image pairs $< I_B, I'_B >$. A keypoints detector (the SIFT) is employed to extract keypoints from image I_B and their corresponding keypoints from image I'_B. This process involves supervised learning. The formula for the keypoints loss is as follows:

$$L_{KP} = \sum_{x_i \in SIFT(I_B)} |f_{B \to B'}(x_i) - T(x_i)|_1 \quad (8)$$

The keypoints loss is used to improve the accuracy of keypoints prediction by the model, particularly for image pairs with considerable differences in perspective (Fig. 6). This loss function increases the initial prediction accuracy and refinement speed as the system enters Stage 2.

2) Interleaved Drone Controller: From the results of the KADFP model, a filtered key flow and its corresponding keypoints alignment error $E^{\text{cur}}_{\text{kpa}}$ are derived. The interleaved drone controller (Fig. 7) is then used to adjust the drone's flight. An evaluation is performed to determine whether $E^{\text{cur}}_{\text{kpa}}$ falls below a specific threshold θ, which was set to 5 and

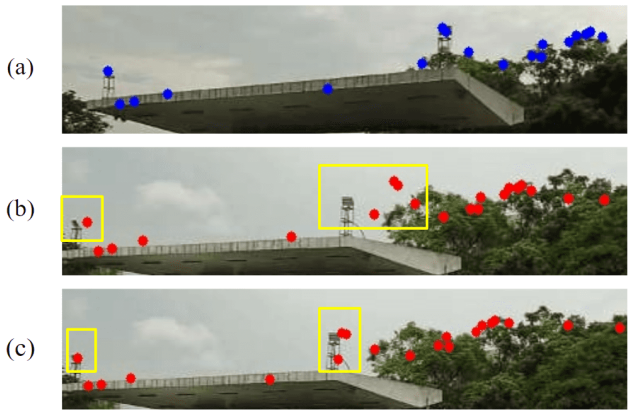


Fig. 6: Effect of the keypoints loss on keypoints prediction: (a) keypoints in the current drone view, and keypoints in target image filtered from the predicted dense flow obtained (b) without using the keypoints loss and (c) using the keypoints loss.

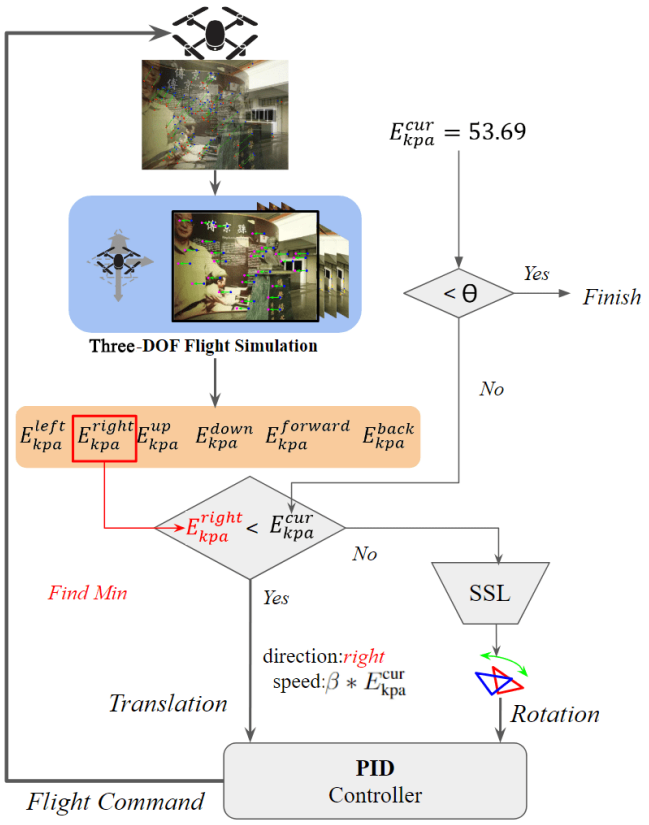


Fig. 7: Architecture of the interleaved drone controller.

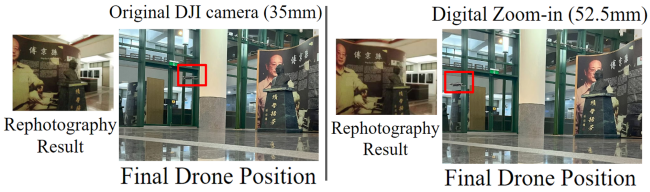


Fig. 8: Different focal lengths of the camera.

10 pixels for indoor and outdoor scenes, respectively, in this study. If the aforementioned parameter is below the threshold, the drone has achieved its target rephotography position; otherwise, the system must fine-tune the drone positioning.

In the position fine-tuning process, the **Three-DOF flight simulation** module is used to simulate the translational behavior of a drone along the x, y, and z axes, covering a total of six scenarios: up, down, left, right, forward, and backward. This simulation is achieved by shifting the keypoints in the opposite direction of the intended drone movement. For instance, to simulate the drone moving upward, keypoints are displaced downward. For movements to the left or right, keypoints are shifted to the right or left, respectively. When forward flight and backward flight are simulated, keypoints are moved away from and toward the image center, respectively. After this simulation, the keypoints alignment error for each direction is computed (labeled as $E_{\text{kpa}}^{\text{direction}}$). The direction of the next flight command is the direction with the smallest error value, as expressed in the following equation:

$$E_{\text{kpa}}^{\text{direction}} = \frac{1}{|K|} \sum_{(X_{\text{drone}}, X_{\text{target}}) \in K} d(X_{\text{drone}} + \Delta X^{\text{direction}}, X_{\text{target}}) \tag{9}$$

The formulas for $\Delta X^{\text{direction}}$ are expressed in (10) and (11), where $Shift$ represents the extent of displacement. In this study, α' is set as 0.1, distinguish from α in (2).

$$Shift = \alpha' * E_{\text{kpa}}^{\text{cur}} \tag{10}$$

$$\Delta X^{\text{direction}} \begin{cases} \Delta X^{\text{forward}} = -Shift * \frac{(X_{\text{drone}} - X_{\text{center}})}{\|X_{\text{drone}} - X_{\text{center}}\|} \\ \Delta X^{\text{backward}} = Shift * \frac{(X_{\text{drone}} - X_{\text{center}})}{\|X_{\text{drone}} - X_{\text{center}}\|} \\ \Delta X^{\text{right}} = (-Shift, 0) \\ \Delta X^{\text{left}} = (Shift, 0) \\ \Delta X^{\text{up}} = (0, Shift) \\ \Delta X^{\text{down}} = (0, -Shift) \end{cases} \tag{11}$$

The smallest $E_{\text{kpa}}^{\text{direction}}$ value is compared with the value of $E_{\text{kpa}}^{\text{cur}}$. If the smallest $E_{\text{kpa}}^{\text{direction}}$ value is smaller than the value of $E_{\text{kpa}}^{\text{cur}}$, then moving the drone in the relevant direction would be the most efficient strategy for matching the current view with the target image's keypoints. The drone then moves in the selected direction at a speed governed by $\beta * E_{\text{kpa}}^{\text{cur}}$, where β can be set to 0.01.

If no directional adjustments yield error reduction, the visual localization method of Stage 1 is executed by focusing solely on rotational adjustments. The instructions of the interleaved drone controller are then processed using the PID controller [26]. Our fine-tuning approach that involves alternate translations and rotations consistently minimizes errors $E_{\text{kpa}}^{\text{cur}}$.

TABLE I: Average $E_{\text{kpa}}^{\text{cur}}$ values for the three scenes.

Scene	Stage1	Our(total)
Indoor (12 times test average)	35.56	4.72
Outdoor1 (7 times test average)	48.92	9.43
Outdoor2 (5 times test average)	60.72	9.62

IV. EXPERIMENTS

A. Implementation Details

We employed a pretrained L-Twins-SVT structure as the image encoder for the second stage of the KADFP model. To conduct dense flow estimation with the KADFP model, we utilized two 3090 Ti graphics processing units and trained the model on the MegaDepth [31] data set. This data set contains more than 13,000 image pairs. We trained our model for 10 epochs, and the experimental scenes encompassed an indoor setting (area of approximately 8 m x 16 m) and two outdoor settings (areas of 10 m x 45 m and 40 m x 50 m). The drone adopted in this study was DJI Mavic Pro.

B. Rephotography Results

Multiple tests were conducted in the three experimental scenes, and the average $E_{\text{kpa}}^{\text{cur}}$ values were calculated; these values are presented in Tab. I. The stage 1 of visual localization results were recorded and then compared with the results obtained after Stage 2. Tab. I indicates that the fine-tuning performed in Stage 2 considerably reduced the $E_{\text{kpa}}^{\text{cur}}$ value. For the indoor scene and outdoor scenes, the error was maintained below 5 and 10 pixels, respectively. As presented in Tab. I, a notable difference existed between the results obtained in Stage 1 for the two outdoor scenes. This finding suggests that the size of the scene affects the visual localization results and thus the accuracy of camera pose estimation. Fig. 9 depicts the rephotography results for the three scenes. The adopted method exhibited high performance even under variable lighting conditions. Fig. 8 demonstrates that even when the focal length differs from that of the target image, our method can still achieve precise rephotography performance by maneuvering the drone. The detailed rephotography process can be viewed in the accompanying demo video: https://www.youtube.com/watch?v=0CVtIfGV5-0.

C. Ablation Study

From the analysis of data in Tab. II, which pertains to indoor scene, we made several comparisons:

L-Twins-SVT versus CNN. L-Twins-SVT, which is a Transformer-based model, excels in extracting features from a pair of images, as indicated by the average $E_{\text{kpa}}^{\text{cur}}$ values in the rephotography results.

With and without the keypoints loss: Calculating the keypoints loss is crucial, especially when considerable visual differences exist between images. Fig. 10 displays the rephotography processes obtained using the KADFP model

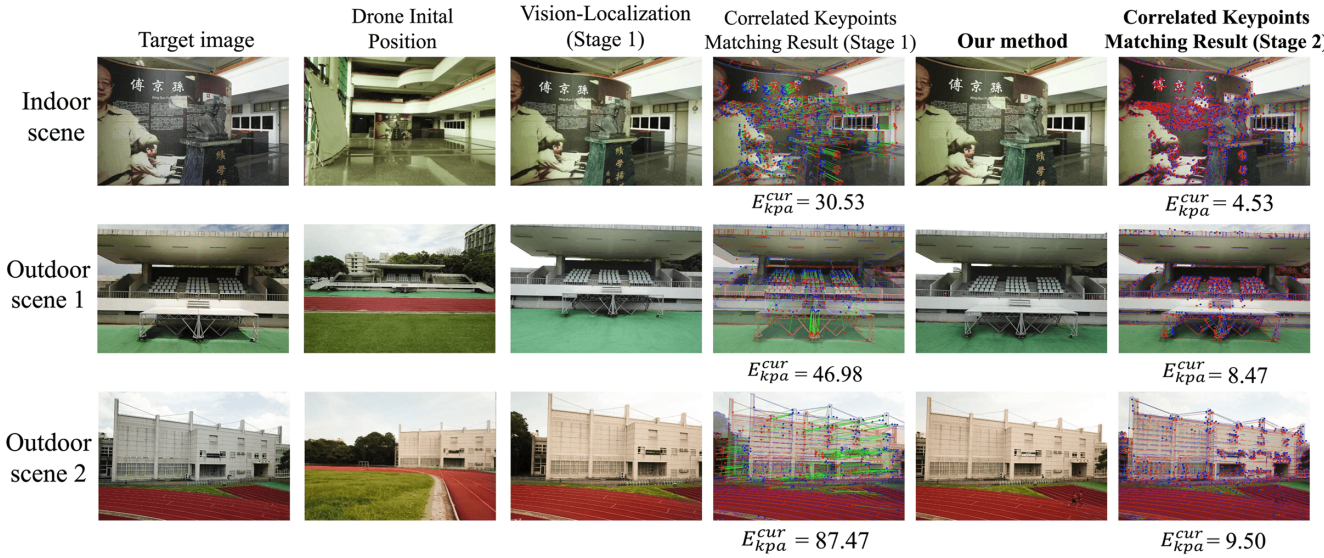


Fig. 9: Rephotography results for the three scenes are presented alongside keypoints matching results. Edges on the blended images show the target image in red and the current drone view in blue, offering a clear view of the rephotography outcome.

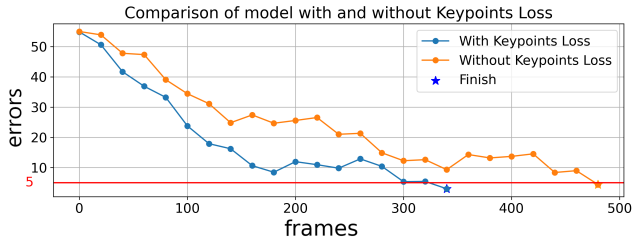
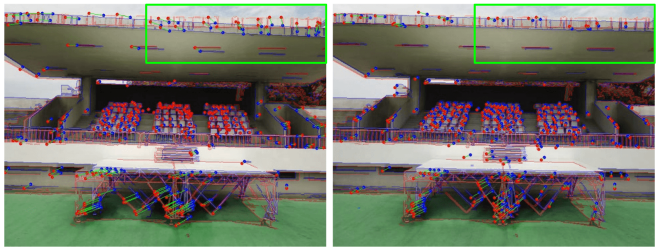


Fig. 10: Results obtained with and without fine-tuning based on the keypoints loss.



(a) (b)

Fig. 11: Comparison of interleaved versus noninterleaved adjustments. (a) using only translation, and (b) using interleaved translation and rotation. The center aligns well in (a), but the upper section of building misaligns due to rotational differences.

with and without the keypoints loss. The initial error was approximately 50; however, the addition of the keypoints loss improved keypoints predictions and accelerated the early fine-tuning stages.

Interleaved versus noninterleaved adjustments: We compared the results obtained with interleaved and noninterleaved adjustments during the fine-tuning of the translation process (Tab. II). Compared with the rephotography results obtained through noninterleaved adjustments, those achieved through interleaved adjustments were notably superior. The green box in Fig. 11 indicates that achieving optimal results purely through translation is unfeasible when discrepancies in angular rotation exist.

TABLE II: Results obtained in the ablation study.

Configuration	$E_{\mathbf{kpa}}^{\mathbf{cur}}$
CNN-encoder	7.82
Without Keypoints loss	4.81
Without Interleaved Drone Controller	6.53
Our Method	**4.72**

V. CONCLUSION

This study proposes a drone rephotography system that integrates computer-vision-based localization and pixel-level dense flow prediction fine-tuning to achieve precise rephotography images with viewpoints that are closely aligned with those of target images. The proposed KADFP model effectively handles challenges such as lighting variations and background differences. Moreover, a novel flight procedure is implemented in the proposed system. This procedure involves using an interleaved drone controller to alternate between translation and rotation adjustments for ensuring smooth flight dynamics during rephotography. The results of this study indicate that the developed system can be used for achieving accurate drone rephotography.

ACKNOWLEDGEMENT

This work was supported in part by Chunghwa Telecom Laboratories, Taoyuan, Taiwan and the National Science and Technology Council, Taiwan (111-2628-E-A49-003-MY2, 112-2634-F-A49-007-, and 113-2221-E-A49-164-MY3).

REFERENCES

[1] Hirano, Y., Garcia, C., Sukthankar, R., Hoogs, A. (2006). Industry and Object Recognition: Applications, Applied Research and Challenges. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds) Toward Category-Level Object Recognition. Lecture Notes in Computer Science, vol 4170. Springer, Berlin, Heidelberg.

[2] J. M. D. Barros, B. Mirbach, F. Garcia, K. Varanasi and D. Stricker, "Fusion of Keypoint Tracking and Facial Landmark Detection for Real-Time Head Pose Estimation," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 2018, pp. 2028-2037, doi: 10.1109/WACV.2018.00224.

[3] Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C. (2014). Learning 6D Object Pose Estimation Using 3D Object Coordinates. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8690. Springer, Cham.

[4] M. P. Muresan, M. Raul, S. Nedevschi and R. Danescu, "Stereo and Mono Depth Estimation Fusion for an Improved and Fault Tolerant 3D Reconstruction," 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2021, pp. 233-240, doi: 10.1109/ICCP53602.2021.9733702.

[5] Image Registration Techniques: An overview. Medha V. Wyawahare, Dr. Pradeep M. Patil, and Hemant K. Abhyankar. International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.2, No.3, September 2009.

[6] G. Balamurugan, J. Valarmathi and V. P. S. Naidu, "Survey on UAV navigation in GPS denied environments," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 2016, pp. 198-204, doi: 10.1109/SCOPES.2016.7955787.

[7] N. Zhu, J. Marais, D. Bétaille, M. Berbineau, GNSS position integrity in urban environments: A review of literature, IEEE Trans. Intell. Transp. Syst. PP (99) (2018) 1–17, http://dx.doi.org/10.1109/TITS.2017.2766768.

[8] From Coarse to Fine: Robust Hierarchical Localization at Large Scale. Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, Marcin Dymczyk. arXiv:1812.03506 [cs.CV] (or arXiv:1812.03506v2 [cs.CV] for this version)https://doi.org/10.48550/arXiv.1812.03506

[9] Pauline C. Ng , Steven Henikoff, SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Research, Volume 31, Issue 13, 1 July 2003, Pages 3812–3814, https://doi.org/10.1093/nar/gkg509

[10] S. S. Beauchemin and J. L. Barron. 1995. The computation of optical flow. ACM Comput. Surv. 27, 3 (Sept. 1995), 433–466. https://doi.org/10.1145/212094.212141

[11] Zachary Teed and Jia Deng. 2020. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, 402–419.

[12] Huang, Z., "LIFE: Lighting Invariant Flow Estimation", ¡i¿arXiv e-prints¡/i¿, 2021. doi:10.48550/arXiv.2104.03097.

[13] Chu, X., "Twins: Revisiting the Design of Spatial Attention in Vision Transformers", ¡i¿arXiv e-prints¡/i¿, 2021. doi:10.48550/arXiv.2104.13840.

[14] M. Jun, S.I. Roumeliotis, G.S. Sukhatme, State estimation of an autonomous helicopter using Kalman filtering, in: Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289), Vol. 3, 1999, pp. 1346–1353, http://dx.doi. org/10.1109/IROS.1999.811667.

[15] Y. Oshman, I. Shaviv, Optimal tuning of a Kalman filter using genetic algorithms, in: AIAA Guidance, Navigation, and Control Conference and Exhibit, 2000, p. 4558, http://dx.doi.org/10.2514/6.2000-4558.

[16] J. Sasiadek, Q. Wang, R. Johnson, L. Sun, J. Zalewski, UAV navigation based on parallel extended Kalman filter, in: AIAA Guidance, Navigation, and Control Conference and Exhibit, 2000, p. 4165, http://dx.doi.org/10.2514/ 6.2000-4165.

[17] B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, GNSS – Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More, Springer-Verlag, Wien, 2008.

[18] H. Kuusniemi, G. Lachapelle, GNSS signal reliability testing in urban and indoor environments, in: Proceedings of the NTM Conference, 2004.

[19] G. Lachapelle, GNSS indoor location technologies, J. Glob. Position. Syst. 3 (1,2) (2004) 2–11, http://dx.doi.org/10.5081/jgps.3.1.2.

[20] G.J. Van Dalen, D.P. Magree, E.N. Johnson, Absolute localization using image alignment and particle filtering, in: AIAA Guidance, Navigation, and Control Conference, 2016, p. 0647, http://dx.doi.org/10.2514/6.2016-0647.

[21] A. Yol, B. Delabarre, A. Dame, J.-E. Dartois, E. Marchand, Vision-based absolute localization for unmanned aerial vehicles, in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 3429–3434, http://dx.doi.org/10.1109/IROS.2014.6943040.

[22] J. Lewis, Fast template matching, Vis. Interface 95 (1994) 15–19.

[23] C. Harris, M. Stephens, A combined corner and edge detector, in: Procedings of the Alvey Vision Conference 1988, Alvey Vision Club, Manchester, 1988, pp. 23.1–23.6.

[24] A. Marcu, D. Costea, E. Slusanschi, M. Leordeanu, A multi-stage multi-task neural network for aerial scene interpretation and geolocalization, 2018, arXiv:1804.01322 [cs].

[25] M. Schleiss, Translating aerial images into street-map representations for visual self-localization of uavs, ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 4213 (2019) 575–580, http://dx.doi.org/10.5194/isprsarchives- XLII-2-W13-575-2019.

[26] WILLIS, M. J. Proportional-integral-derivative control. Dept. of Chemical and Process Engineering University of Newcastle, 1999, 6.

[27] J. L. Schönberger and J. -M. Frahm, "Structure-from-Motion Revisited," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 4104-4113, doi: 10.1109/CVPR.2016.445.

[28] LU, Xiao Xin. A review of solutions for perspective-n-point problem in camera pose estimation. In: Journal of Physics: Conference Series. IOP Publishing, 2018. p. 052009.

[29] CANTZLER, H. Random sample consensus (ransac). Institute for Perception, Action and Behaviour, Division of Informatics, University of Edinburgh, 1981, 3.

[30] DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. "Superpoint: Self-supervised interest point detection and description." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018.

[31] MegaDepth: Learning Single-View Depth Prediction from Internet Photos Zhengqi Li, Noah Snavely Cornell University/Cornell Tech, In CVPR, 2018

[32] H. Cai, S. Ye, A. Vardy and M. Gong, "3D Visual Homing for Commodity UAVs," 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 2018, pp. 269-276, doi: 10.1109/CRV.2018.00045.

[33] Damian M. Lyons and Noah Petzinger "Visual homing for coordinated robot team missions", Proc. SPIE 12549, Unmanned Systems Technology XXV, 1254902 (14 June 2023); https://doi.org/10.1117/12.2664413

[34] Chen, K. W., Xie, M. R., Chen, Y. M., Chu, Lin, Y-B. (2022). DroneTalk: An Internet-of-Things-Based Drone System for Last-Mile Drone Delivery. IEEE Transactions on Intelligent Transportation Systems, 23(9), 15204 . [9703276]. https://doi.org/10.1109/TITS.2021.3138432