# Hierarchically Organized Computer Vision in Support of Multi-Faceted Search for Missing Persons

Arturo Miguel Russell Bernal and Jane Cleland-Huang Computer Science and Engineering Department, University of Notre Dame, Indiana, USA arussel8@nd.edu, JaneHuang@nd.edu

Abstract-Missing person searches are typically initiated with a description of a person that includes their age, race, clothing, and gender, possibly supported by a photo. Unmanned Aerial Systems (sUAS) imbued with Computer Vision (CV) capabilities, can be deployed to quickly search an area to find the missing person; however, the search task is far more difficult when a crowd of people is present, and only the person described in the missing person report must be identified. It is particularly challenging to perform this task on the potentially limited resources of an sUAS. We therefore propose AirSight, as a new model that hierarchically combines multiple CV models, exploits both onboard and off-board computing capabilities, and engages humans interactively in the search. For illustrative purposes, we use AirSight to show how a person's image, extracted from an aerial video can be matched to a basic description of the person. Finally, as a work-in-progress paper, we describe ongoing efforts in building an aerial dataset of partially occluded people and physically deploying AirSight on our sUAS.

#### I. Introduction

The deployment of Unmanned Aerial Systems (sUAS) is becoming increasingly prevalent in diverse application areas such as smart home construction [21], traffic and fire surveillance, environmental sampling, law-enforcement activities [13], and search and rescue missions [2]. In such scenarios, quickly finding a human in need or identifying a suspect in a crowd can save lives. Currently, emergency responders tend to fly sUAS manually to enhance operations by providing additional aerial, mobile "eyes"; however, autonomous sUAS could be even more beneficial in time-critical scenarios [9], empowering human first responders and law enforcement officers to focus on mission level goals supported by intelligent sUAS [2], [9].

An sUAS's onboard computer vision (CV) system plays a critical role in enabling its autonomy. However, deploying an effective CV pipeline on a resource-constrained, edge-based computing platform is very challenging. Edge computing forces the system to perform operations under severely *limited resources* [17], [22], which can degrade performance, result in loss of accuracy, and potentially lead to failure to detect a human even when they appear in the image stream. We address this challenge through proposing *AirSight* – an *optimizable*, *hierarchical approach* that seeks to improve the performance and accuracy of sUAS-based CV search for a missing person. *AirSight* assumes an sUAS infrastructure that includes both *on-board* (onboard the Unmanned Aerial

Vehicle) and *off-board* (Ground Control Station) computing capabilities. Initial vision tasks are executed *on-board* to identify *objects of interest*, and then the remainder of the CV task is delegated to the resource rich, ground-based infrastructure, where humans can also be engaged as active decision makers [2], [9].

The remainder of this paper is structured as follows. Section II presents our hierarchical approach to CV-based person detection, while Section III describes a simple case study that illustrates the use of our approach. Section IV describes related work that uses both CV and non CV approaches to aid aerial detection. Finally, Sections V and VI describe future work, draw conclusions, and provide a collaboration statement.

#### II. A CONFIGURABLE HIERARCHICAL APPROACH

AirSight addresses the issue of limited resources for sUAS-based edge CV tasks, such as detecting and identifying a missing person, by exploiting and configuring various combinations of CV models, and deploying them across on-board and off-board computing resources. As illustrated in Fig. 1, when applied to the task of finding a missing person, the mission starts when (1) a description of the missing person is provided and the sUAS are prepared for deployment. Next, (2) the CV hierarchy is configured by (3) deploying lightweight CV models on-board for initial person detection and filtering, and (4) deploying additional models off-board for further analysis with (5) support for human-in-the-loop decision making.

Each sUAS must be equipped with a camera, gimbal, and a CV pipeline. Installing this pipeline on the edge device is very challenging due to several tradeoffs. The computer needs to provide enough storage and memory to install and run CV algorithms, models, and other flight-related software, but also be as light as possible to increase flight time. In addition, the onboard computer must be sufficiently powered with adequate ventilation and cooling to avoid problems such as cutting-out or throttling down during flight.

In order to differentiate between different people, *Air-Sight* includes traditional CV models for detecting people and differentiating characteristics such as gender and age, while using heuristic-driven approaches for detecting clothing color and other characteristics. *AirSight* runs a basic persondetection algorithm on-board, and one or more additional CV models to perform an initial evaluation as to whether each detected person fits key aspects of the search criteria.

However, the limited computational resources available onboard the sUAS mean that only a subset of models can be executed on-board the sUAS (*On-Board computing*), while the remainder need to run off-board on the Ground Control Station (GCS) (*Off-Board computing*). Furthermore, it makes sense to deploy discriminatory models on-board in order to quickly filter out unlikely persons. All non-filtered images are sent to the GCS for further analysis by other CV models. Overall, our off-board approach represents a traditional Blackboard-style architecture in which various agents (CV algorithms and models) contribute their opinion about the extent to which the detected person matches the provided description [5]. Any person scoring above a predefined likelihood threshold can then be presented to the human for evaluation.

## A. Mission startup

The mission starts when an alert is raised for a missing person and a description and/or photograph is provided. The emergency responders quickly activate and deploy a cohort of CV-equipped sUASs charged with the task of person detection and identification. During this step, the human team member shares the description of the missing person with the sUAS in a machine-readable format (e.g., JSON), with additional information such as the region of the search, that could be used to quickly plan the mission.

## B. Hierarchical configuration

This step defines which CV models are to be included in the analysis as well as their hierarchical arrangement. Table I provides a non-exhaustive list of the types of elements that could be included in a description of a missing person. A CV model is included in the hierarchy for each key characteristic included in the missing person description.

Table II, represents a configuration file that provides a mapping between types of features included in the description, the appropriate CV models to use, and the arrangement of each of them in our hierarchical model. For each particular descriptive attribute, the system can have more than one model, and the position on the hierarchy may determine which model to use. In our current implementation, we manually create the JSON specification which defines the hierarchy of CV models, and determines which models reside on-board or off-board. However, in future designs we plan for this configuration to be performed automatically, as discussed in future work (see Section V).

On-Board resources must be processing light, while Off-Board models have access to more powerful computational resources; therefore, the configuration file indicates the particular model to use and the order in which each CV model is applied. This is depicted in Fig. 1, where models assigned with the same sub-level number are executed in parallel, and models with different sub-level numbers are executed in series, allowing us to include series and parallel sub-levels if needed. Finally, as seen in Table II, further parameters, such as a tolerance for the age range inference, are also specified.

#### TABLE I

Sample json file with the information of a missing person visual attributes.

```
"clothing_color": [str, str],
"clothing": [str, str],
"age": int,
"gender": str,
"race": str,
"skin_tone": str,
"reference_image": path_str,
"height": int,
"weight": int,
"special": [str, str, ...]
```

TABLE II

CONFIGURATION OF THE CV MODELS SPECIFIED IN A JSON FILE.

```
"compute_onboard": {
    "filter-human": {"model": str, "sublevel": int},
    "filter-clothing_color": {"model": str, "sublevel": int},
    ...
},
"compute_offboard": {
    "filter-clothing_color": {"model": str, "sublevel": int},
    "filter-age": {
        "model": str, "sublevel": int, "tolerance": int},
    ...
},
"compute_human": int
```

AirSight deploys a lightweight discriminatory model onboard the sUAS to serve as a filtering mechanism. For example if the missing person was wearing a red shirt with blue shorts, then a clothing-based CV model could provide strong differentiation and any person not wearing at least one of these items should be filtered out. However, as pointed out by emergency responders whom we interviewed, in the case of lost (and potentially abducted) children, the value of clothing descriptions dissipates quickly as the kidnapper could change the child's clothes. In this case, the onboard model might filter by body size or by age instead of clothing, and additional mission-specific attributes such as shoes description could be deployed for off-board computing.

Fig. 1 shows the elements of our model as well as their hierarchy and order of execution. It includes the following:

- Three primary levels representing onboard, groundbased, and human analysis.
- Sub-levels representing a group of *filter-models* that are separated from each other by a *discriminatory logic phase*.
- Filter-models that address the detection/inference of a particular attribute of the missing person, and can be grouped to act in parallel and make a joint decision using a blackboard style approach, or placed in a series to create a pipeline, as observed in Fig. 1.

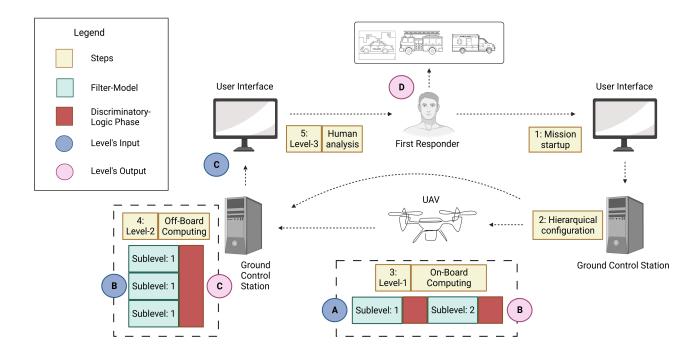


Fig. 1. The Hierarchical CV model is composed of three different levels, each one including a set of filters executed either as a pipeline or in parallel supported by a blackboard style architecture. The filters at each level are responsible for determining whether an image should be passed for further analysis to the next layer. The final layer includes human inspection.

## C. Level-1: On-Board computing

In our adopted architecture the on-board computer is responsible for performing the first image processing steps. Fig. 1 shows a series configuration where the first CV model is responsible for processing the raw frames provided by the video stream, followed by a *discriminatory logic phase* in which it determines whether the analyzed frames should be passed to the next sub-level for further processing. Rather than passing each and every frame, the filter can also identify representative, high-quality frames representing the highest likelihood of matching the missing person description. Level-1 is responsible for decreasing the amount of information that is passed to the GCS. Its role is highly important for reducing network traffic whilst optimizing the likelihood of finding the targeted person.

## D. Level-2: Off-Board computing

After receiving the data from Level-1, the off-board computer is responsible for detecting and/or inferring the attributes of a missing person that were not analyzed in Level-1 due to limited resources. However, it could also apply more powerful models to the same attributes addressed by Level-1, or even include facial recognition algorithms, with the goal of reducing false positives. Fig. 1 illustrates a parallel configuration on Level-2, where three models process the same information provided by Level-1, either addressing the same attributes with different computer vision models in order to improve confidence of the result, or addressing different attributes of interest.

## E. Level-3: Human analysis

Finally, the candidate video streams or images that passed through the discriminatory logic phases of Level-1 and Level-2 are shown to a human team member monitoring and supervising the sUASs, who provides feedback to *Air-Sight* about why the person does not match the description, and/or dispatches physical resources such as additional sUASs, other equipment, or human team members, to verify the accuracy of the candidate.

The design of the interface between Level-2 and Level-3 is still to be explored, one option being that the human team member has the ability to choose the format of the user interface, the amount of information displayed, and even to control the level of autonomy of the sUAS. A space is placed on the configuration file, shown in Table II, that could allow us to configure the hierarchy model for Level-3, with modifications to the user interface. By integrating the human as a team member, we are seeking for an improvement in performance not only by the permutation of computer vision models, but by the communication and integration of the human team member in the sUAS pipeline.

#### III. CASE STUDY

We conducted a case study which serves as a preliminary proof-of-concept for *AirSight*. Fig. 2 shows three levels of the pipeline. The images used for the case study were extracted from an on-going IRB approved dataset collection that we are performing, which currently includes 75 subjects out of a planned 100 subjects. The dataset includes aerial views of humans in various poses and varying degrees of occlusion

behind diverse obstacles. When completed, we will publicly release the dataset, with annotated aerial imagery of fully visible and occluded subjects as well as a 360 degree low altitude view of each subject to serve as a reference image. For purposes of this preliminary case study, we selected 12 subjects from the collection using a stratified approach which resulted in balanced gender (six males and six females), age (six adults/older adults and six young persons), and skin tone (six white and six non-white subjects). We extracted five frames of whole-body images for each subject to create a mini-dataset of 60 frames that was then used in this study.

# A. Start of Mission

Out of the 12 subjects we selected one to serve as our missing person, and used the demographic data associated with the subject to form a missing person description, with age (28 years old), gender (male), race (latino hispanic), and clothing color (blue shirt and black pants). This is summarized in Table III. It should be noted that we deliberately selected a person wearing easily described clothing and not patterned material for this proof-of-concept. The case study simulates the situation of a missing person with a description but not a photograph. In addition, the case study provides an example in which the person is expected to be wearing the described clothing (e.g., immediately after a person goes missing, or a shooting suspect who has escaped from the vicinity of the crime).

#### B. Hierarchical configuration

As no picture is provided in our scenario, searching using facial recognition is not possible. Furthermore, even if a picture was available from the ground, the perspective from an sUAS could be very different, therefore in this example we focus on the textual description. The JSON file describing the person is shown in Table III, and a second JSON file describing the hierarchy of CV models is provided in Table IV. These files are sent to both the sUAS and to the GCS service that performs off-board CV. Each of these activates the relevant CV models and establishes a pipeline as specified in the JSON file. The specification dictates exactly when frames will be passed from one model to another, and when they will be transmitted from the sUAS to the ground-based service, and from the ground-based CV service to the human.

In our case study we utilize clothing color as the preliminary filter on-board the sUAS and assign other CV models for age, gender, and race to the off-board GCS.

## C. Level-1: On-Board computing

Fig. 2 shows the deployment of the hierarchical CV process. The YOLOv4 model was used to detect a person using a threshold confidence score. YOLOv4 was able to successfully detect a human in 59 of the 60 images, based on our predefined confidence score of 0.60. Bounding boxes for the 60 images were manually annotated in order to evaluate the performance of the YOLOv4 model on these images, using an Intersection over Union (IOU) value of 50%. Table V shows the results when varying the confidence threshold.



Fig. 2. The image shows the pipeline of actions for a case study of the hierarchical model through the three different levels.

Images of detected persons were cropped, and all subsequent analysis was performed on the cropped images.

For illustrative purposes we show how one frame, taken from the targeted subject, was analyzed using the CV hierarchy. In addition to the person detector we also deployed a second CV algorithm on-board the sUAS with the aim of checking whether the detected person's clothing matched the description. The color-clothing filter-model used a region-based segmentation [18] that provided a segmentation of foreground and background.

The assumption taken during this part of the case study was that the first sub-level would provide an image of a standing whole body person, therefore a histogram could be applied to the upper and lower part of the segmented foreground image. The histogram was analyzed in the HSV color space, where most common colors could be distinguished just by the hue component. Besides simplifying the classification of colors, this approach introduces robustness against different levels of light over the same color. Once the histogram was built, the model looked for the peak values and reported an HSV value of [228,80,47] for the top of the image and an HSV value of [240,33,4] for the bottom of the image. These values were compared against a preset list of HSV values of common colors and were reported back as blue and black. It should be noted that the classification of black, grey and white in the HSV color space are based on the saturation and value numbers, ignoring the hue number. Finally, the discrimination logic phase applied during this part was a complete match, that is, the upper-body color that was detected matched the t-shirt color provided in the description, while the lower-part color matched the pants color; therefore this cropped image was passed to Level-2. We discuss challenges of using clothing colors in Section V.

# D. Level-2: Off-Board computing

In our case study example, Level-2 used a parallel configuration to analyze gender, age and race. The DeepFace

"cl	othing_color": ["blue", "black"],		
"aş	ge": 28,		
"ge	"gender": "male",		
"ra	ce": "latino hispanic"		

TABLE IV

Sample json file of the configuration of the computer vision models for the case study.

```
"compute_onboard": {
    "filter-human": {
        "model": "YOLOv4_pretrained", "sublevel": 1},
    "filter-clothing_color": {
        "model": "foreground_seg", "sublevel": 2},
},
"compute_offboard": {
        "filter-age": {
            "model": "deepface", "sublevel": 1, "tolerance": 5},
        "filter-gender": {"model": "deepface", "sublevel": 1},
        "filter-race": {"model": "deepface", "sublevel": 1}
},
"compute_human": int
```

library was used as it incorporates state-of-the-art models for face detection, face verification and face recognition tasks, and includes facial attribute analysis of age, gender, race and facial expressions [15], [16]. The DeepFace facial attribute analysis model has reported a Mean Average Error (MAE) of +-4.6 years for age inference, and a 97.44% accuracy, 96.29% precision and 95.05% recall for gender inference [16]. An SSD face detection model was used from the DeepFace library to crop the face from the whole body image. DeepFace provides access to models with better face alignment such as RetinaFace and MTCNN if required. After having the face image of the subject, this image was fed to the DeepFace facial attribute analysis model. Table VI shows a comparison of the ground truth and the results obtained from DeepFace. Although race was incorrectly inferred, the discrimination logic phase used in this sub-level employed a simple majority voting scheme, meaning that if two of the attributes matched the description, the candidate image would be passed to the human in Level-3. Detected age was deemed a match as it was within the tolerance of +-5 years set in the configuration file.

# E. Level-3: Human Team Member

Finally, images that passed the criteria assessed in Level-2, were presented to a human team member, who then applied their own discrimination logic to determine whether the image represented the missing person or not.

# IV. RELATED WORK

The interdisciplinary nature of our work means that our solution leverages ideas from software engineering, human

 $\label{thm:table V} TABLE\ V$  Results of the YOLOv4 model over the 60 sample images.

Confidence threshold	TP	FP	FN	Recall	Precision
0.00	59	42	1	0.98	0.58
0.60	59	24	1	0.98	0.71
0.80	56	10	4	0.93	0.84
0.85	51	3	9	0.85	0.94
0.90	46	0	14	0.76	1.00

TABLE VI INFERENCE RESULTS OBTAINED FROM DEEPFACE

Attribute	Truth	Inference			
Age[years]	28	23			
Gender	Male	Man			
Race	Latino Hispanic	asian: 1.0718737728893757,			
		indian: 3.075053170323372,			
		black: 0.10856959270313382,			
		white: 57.12805986404419,			
		middle eastern: 24.243606626987457,			
		latino hispanic: 14.372840523719788			

computer interaction, hardware development and computer vision in order to address an open challenge in the field of emergency response scenarios and aerial identification.

## A. Computer vision approaches

Several researchers have developed new datasets, such as DroneFace [10] and DroneSURF [11] for aerial face identification, VisDrone [20] for aerial object detection and tracking, and HERIDAL [12] and SARD [14] for aerial human detection under search and rescue missions. A more extensive description of each dataset, as well as additional categorized datasets, can be found in [19]. These datasets have enabled CV algorithms to be trained specifically for aerial detection and identification tasks, and have enhanced the ability of researchers to develop improved CV models for aerial tasks. Amato et al. [4] address the low resolution characteristic of drone imagery, and test an optimized CV model for cross-resolution face re-identification (gallery images with high resolution vs probe images of low resolution) on the dataset DroneSURF, achieving an accuracy of 60.87%. Despite the improved accuracy, further improvements are needed to support edge-based tasks such as identification of a missing person. Abraham et al. [1] seek improvement of CV models performance during aerial human detection on emergency response scenarios by taking into consideration the context of the detection, resulting in decreasing the number of false positives. Božić-Štulić et al. [6] introduce a new region proposal network algorithm, inspired by the Faster R-CNN model, but using saliency object detection algorithms for the Region Proposal Module; their tests on the HERIDAL dataset acheive a detection rate of 88.9%.

Our *AirSight* work proposes an optimizable hierarchical architecture that exploits existing CV models to provide an effective CV pipeline for person detection and identification in a limited resource environment. This differs from the majority or prior work, which has emphasized detection rather than identification (cf. Fig.3) [19].

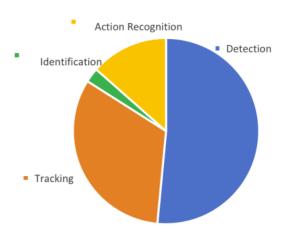


Fig. 3. Aerial computer vision papers distribution [19].

## B. Non-computer vision approaches

Other research has focused on integrating additional hardware and software capabilities into sUAS for their deployment in search and rescue missions. Pensieri *et al.* [13] explore the idea of introducing remote sensing tools, such as ground penetrating radars, to aid in the search for a missing person. Albanese *et al.* [3] introduce an interesting concept of incorporating mobile networks into the sUAS to localize a cellular network during a search and rescue mission, under the assumption that the missing person carries a cellphone and that the signal detected corresponds to the missing person. Their system also guides the sUAS towards a trajectory that would enhance the localization accuracy. Such solutions could be integrated with our approach to improve the overall mission performance.

Additionally, we have explored areas of Human Computer Interaction (HCI) and software engineering. Agrawal *et al.* [2], through participatory design with first responders, designed a solution for achieving situational awareness during the deployment of multi-sUAS for emergency response applications. Further research aimed at tackling different types of interactions in a human-sUAS environment [7]. Furthermore, Cleland-Huang *et al.* [9] promote partnerships between humans and sUAS by incorporating the Human-Machine Teaming (HMT) paradigm design into the MAPE-K feedback loop model for self-adaptive and autonomous machines, emphasizing interactions, partnerships, and teamwork between humans and machines, capitalizing upon the respective strengths of both the human and the machine, and compensating for each of their potential limitations.

Our approach contributes to the overall engineered solution for the case of search for a missing person by focusing on engineering the computer vision pipeline. In addition, HCI components (described as future work on Section V) address the particular interaction interface needed between emergency responders and sUAS when using our hierarchical approach to search for a missing person.

#### V. FUTURE WORK

In this work-in-progress paper we have proposed *Air-Sight* as a hierarchical approach to supporting multi-faceted, CV-based aerial search for missing persons. Our approach took into consideration the limited resources that are available onboard sUAS due to trade-offs related to weight and power consumption. Our case study demonstrated that our approach is viable; however, the following open challenges, risks, and opportunities should be explored in future work.

- Complex descriptions: In our case study we selected a person wearing clearly defined clothing (blue and black); however, real-world descriptions are often less clear (e.g., strips, patterns, non-descript colors). This served as a proof-of-concept example but clearly oversimplified the real-world. Future work will first evaluate our current approach against our complete ongoing dataset of images, and then explore techniques for improving results for more diverse descriptions and subjects.
- Feature Interactions: Our case study used a very simple majority voting scheme to determine whether to pass a frame to the next level; however, future work will explore more sophisticated approaches for integrating diverse sources of evidence in order to improve the accuracy of the overall approach and to better prioritize the way in which images are forwarded to the next layer of the hierarchy.
- Optimizing CV Hierarchies: In this work we selected an organization of CV models somewhat subjectively; however, future work will comparatively evaluate alternative ways of connecting and arranging CV models. Furthermore, we plan to automate the deployment of AirSight into our own DroneResponse system, which already has fully deployed and tested capabilities for onboard CV supported by off-board microservices capable of running powerful CV algorithms [9].
- **Human-in-the-loop:** Future work will explore ways in which humans can partner with sUAS in person search, and develop effective UIs to support these collaborations
- **Temporal evidence:** Finally, our current approach treats each frame as an isolated entity. In future work we will explore ways to build evidence over a sequence of frames, and how to select best frames to process and forward to subsequent CV models.
- Aerial dataset: Our initial CV experiments have shown reduced accuracy when detecting people from the air, especially when they are partially occluded by objects such as trees. We are therefore collecting and annotating a new dataset of partially occluded subjects from various altitudes. We have collected data from 75 out of 100 planned participants and expect to release the dataset in Spring 2023.

# VI. CONCLUSION AND COLLABORATION STATEMENT

As explained earlier in this paper, deploying CV solutions in real-world conditions can be very challenging. This is

especially true onboard an sUAS in a resource-constrained, aerial environment where computing resources and power are limited, and views of people from above may make it difficult to identify specific individuals. Our hierarchical approach overcomes many of these challenges by leveraging onboard CV capabilities to perform an initial image analysis and then passing images off-board for deeper assessment by a diverse set of CV models and by engaging humansin-the-loop. Our approach can be configured according to specific scenarios by dynamically organizing the hierarchy of CV models, allowing different CV models to be used as preliminary on-board filters, and others to provide deeper assessments as off-board CV services.

While we have focused on the CV pipeline, this work represents an interdisciplinary approach combining the fields of computer vision, software and systems engineering, and HCI research. For example, techniques that engage humans in close partnerships with sUAS, require carefully designed GUIs which allow humans to clearly communicate mission goals and to work interactively with the sUAS in evaluating images of candidate missing persons. Finally, while it may seem simple to deploy a CV pipeline onto an edge device, it is challenging to balance diverse computational needs on a resource-constrained edge device, and therefore solutions must be drawn from software architecture and IoT domains. The CV solution proposed in this paper is part of a broader collaborative effort between our DroneResponse [8] research group at the University of Notre Dame, and several Police and Fire Departments. These have included the South Bend Fire Department [2], Michigan City Fire Department, Mishawaka Fire Department, Mishawaka Police Department, and the University of Notre Dame Police Department. In particular, we have conducted focus groups with each of these agencies to elicit their requirements and to engage them in participatory design activities for deploying sUAS in emergency response scenarios.

## ACKNOWLEDGEMENT

The work in this paper was funded by the USA National Science Foundation Grant CPS-1931962. In addition, we thank Luke Siela for his help with some preliminary experiments.

### REFERENCES

- [1] S. J. Abraham, Z. Carmichael, S. Banerjee, R. G. VidalMata, A. Agrawal, M. N. A. Islam, W. J. Scheirer, and J. Cleland-Huang. Adaptive autonomy in human-on-the-loop vision-based robotics systems. In 1st IEEE/ACM Workshop on AI Engineering - Software Engineering for AI, WAIN@ICSE 2021, Madrid, Spain, May 30-31, 2021, pages 113-120. IEEE, 2021.
- [2] A. Agrawal, S. J. Abraham, B. Burger, C. Christine, L. Fraser, J. M. Hoeksema, S. Hwang, E. Travnik, S. Kumar, W. J. Scheirer, J. Cleland-Huang, M. Vierhauser, R. Bauer, and S. Cox. The next generation of human-drone partnerships: Co-designing an emergency response system. In R. Bernhaupt, F. F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjøn, S. Zhao, B. P. Samson, and R. Kocielnik, editors, CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020, pages 1-13. ACM, 2020.
- [3] A. Albanese, V. Sciancalepore, and X. Costa-Pérez. Sardo: An automated search-and-rescue drone-based solution for victims localization. IEEE Transactions on Mobile Computing, 21(9):3312-3325, 2022.

- [4] G. Amato, F. Falchi, C. Gennaro, F. V. Massoli, and C. Vairo. Multiresolution face recognition with drones. In V. Vozenilek, editor, SSIP 2020: 2020 3rd International Conference on Sensors, Signal and Image Processing, Prague, Czech Republic, October 9-11, 2020, pages 13-18. ACM, 2020.
- [5] L. J. Bass, P. C. Clements, and R. Kazman. Software architecture in practice. SEI series in software engineering. Addison-Wesley-Longman, 1999.
- D. Bozic-Stulic, Z. Marusic, and S. Gotovac. Deep learning approach in aerial imagery for supporting land search and rescue missions. Int. I. Comput. Vis., 127(9):1256-1278, 2019.
- [7] J. Cleland-Huang and A. Agrawal. Human-drone interactions with semi-autonomous cohorts of collaborating drones. In M. A. Baytas, M. Funk, S. Ljungblad, J. Garcia, J. L. Delfa, and F. F. Mueller, editors, Proceedings of the Interdisciplinary Workshop on Human-Drone Interaction co-located with the 2020 ACM CHI Conference on Human Factors in Computing Systems (CHI 2020), Honolulu, HI, April 26, 2020, volume 2617 of CEUR Workshop Proceedings. CEUR-WS.org, 2020.
- J. Cleland-Huang, A. Agrawal, M. N. A. Islam, E. Tsai, M. V. Speybroeck, and M. Vierhauser. Requirements-driven configuration of emergency response missions with small aerial vehicles. In R. E. Lopez-Herrejon, editor, SPLC '20: 24th ACM International Systems and Software Product Line Conference, Montreal, Quebec, Canada, October 19-23, 2020, Volume A, pages 26:1-26:12. ACM, 2020.
- J. Cleland-Huang, A. Agrawal, M. Vierhauser, M. Murphy, and M. Prieto. Extending MAPE-K to support human-machine teaming. In B. R. Schmerl, M. Maggio, and J. Cámara, editors, International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2022, Pittsburgh, PA, USA, May 22-24, 2022, pages 120-131. ACM/IEEE, 2022.
- [10] H.-J. Hsu and K.-T. Chen. Droneface: An open dataset for drone research. In Proceedings of the 8th ACM on Multimedia Systems Conference, MMSys'17, page 187-192, New York, NY, USA, 2017. Association for Computing Machinery.
- I. Kalra, M. Singh, S. Nagpal, R. Singh, M. Vatsa, and P. B. Sujit. Dronesurf: Benchmark dataset for drone-based face recognition. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), pages 1-7, 2019.
- [12] Z. Marusec, D. Bozic-Stulic, S. Gotovac, and T. Marusec. Region proposal approach for human detection on aerial imagery. In 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), pages 1-6, 2018.
- [13] M. G. Pensieri, M. Garau, and P. M. Barone. Drones as an integral part of remote sensing technologies to help missing people. Drones, 4(2), 2020.
- [14] S. Sambolek and M. Ivasic-Kos. Automatic person detection in search and rescue operations using deep CNN detectors. IEEE Access, 9:37905-37922, 2021.
- [15] S. I. Serengil and A. Ozpinar. Lightface: A hybrid deep face recognition framework. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pages 1-5, 2020.
- S. I. Serengil and A. Ozpinar. Hyperextended lightface: A facial attribute analysis framework. In 2021 International Conference on
- Engineering and Emerging Technologies (ICEET), pages 1–4, 2021. [17] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. Edge computing: Vision and challenges. IEEE Internet Things J., 3(5):637-646, 2016.
- [18] Skikit-Image. Comparing edge-based and region-based segmentation. https://scikit-image.org/docs/stable/auto\_ examples/applications/plot\_coins\_segmentation. html,Last accessed on 2022-10-27.
- [19] K. N. Thanh, C. Fookes, S. Sridharan, Y. Tian, F. Liu, X. Liu, and A. Ross. The state of aerial surveillance: A survey. CoRR, abs/2201.03080, 2022. [20] Vision Group. Visdrone dataset, 2017. http://aiskyeye.com/,
- Last accessed on 2022-10-27.
- S. Xia, R. Chandrasekaran, Y. Liu, C. Yang, T. S. Rosing, and X. Jiang. A drone-based system for intelligent and autonomous homes. SenSys '21: The 19th ACM Conference on Embedded Networked Sensor Systems, Coimbra, Portugal, November 15 - 17, 2021, pages 349-350. ACM, 2021.
- [22] X. Xie and K. Kim. Source compression with bounded DNN perception loss for iot edge computer vision. In S. A. Brewster, G. Fitzpatrick, A. L. Cox, and V. Kostakos, editors, The 25th Annual International Conference on Mobile Computing and Networking, MobiCom 2019, Los Cabos, Mexico, October 21-25, 2019, pages 47:1-47:16. ACM, 2019.