Human-Swarm Interaction with a Gesture-Controlled Aerial Robot Formation for Safety Monitoring Applications

Giuseppe Silano[1], Vít Krátký[1⋆], Matouš Vrba[1], Christos Papaioannidis[2], Ioannis Mademlis[2], Robert Pěnička[1], Ioannis Pitas[2], and Martin Saska[1]

***Abstract*— This paper presents a formation control approach for contactless Human-Swarm Interaction (HSI) using hand gestures with a team of multi-rotor Unmanned Aerial Vehicles (UAVs). The approach aims to monitor the safety of human workers, especially those working at heights. In the proposed scheme, one UAV acts as the leader of the formation and is equipped with sensors for human worker detection and gesture recognition. The follower UAVs maintain a predetermined formation relative to the worker's position, thereby providing additional perspectives of the monitored scene. The use of hand gestures allows the human worker to specify movements and action commands for the UAV team, without the need for an additional communication channel or specific markers including the relative distance. Field experiments with three UAVs and a human worker in a mock-up scenario showcase the effectiveness and responsiveness of the proposed approach.**

I. Supplementary Material

Video: `http://mrs.felk.cvut.cz/hmri2023gestures`

II. Introduction

In recent years, there has been a growing interest regarding Unmanned Aerial Vehicles (UAVs), especially multi-rotors. These platforms have gained attention due to their agility, maneuverability, and ability to incorporate diverse onboard sensors [1], [2]. Their modular design and versatility make them well-suited for various applications, including contactless interactions [3], physical engagements with the environment [4], wireless communications [5], aerial filming [6], surveillance, and search & rescue missions [7].

Moreover, UAVs have shown to be advantageous especially in difficult-to-access real-word environments, such as work environments at heights [8], wind turbines [9], large construction sites [10], and power transmission lines [11]. These scenarios often require specialized personnel, expensive equipment, and dedicated vehicles. Introducing UAVs as *robotic co-workers* [12] in these contexts brings numerous benefits, including the ability to observe locations hard to reach by a human, assist in tool handling, ensure workers' safety, and alleviate the physical and cognitive workload on operators [13], [14]. However, contactless interaction with such UAVs is crucial for safety of both the operator and the robot and must be considered in the design of these solutions [15].

Fig. 1: Snapshot showing the gesture-based interaction between a human worker and a team of three UAVs.

Despite significant amount of research focused on collaborative and safe interactions involving human and ground robots, studies involving UAVs have been relatively limited [16]. This gap becomes even more prominent when considering the interaction between humans and multi-robot teams, which is referred to as Human-Swarm Interaction (HSI) [17]. Extensive research has been conducted in the fields of *computer vision* and *autonomous systems* to enhance interaction with humans, enable UAVs to navigate autonomously and avoid unsafe behaviors [17], [18]. Computer vision studies have mostly focused on specific scenarios, exploring mechanisms for detecting faces [19], hand gestures [20], and human body postures [21]. Some works have combined computer vision with audio for multi-modal interfaces [22] and investigated gaze detection for robot selection [23]. Autonomous capabilities for UAVs have been explored, including perception-aware control strategies [24], formation control for visibility enhancement [6], [25], and obstacle avoidance through optimization methods [3], [26]. These studies address collision avoidance with humans, other UAVs, and static objects, as well as enhancing autonomy through local mapping [27].

However, these existing works tend to focus primarily on either the vision component, neglecting or oversimplifying the vehicle dynamics, or the control aspect, abstracting the use of generic onboard sensors. Not considering both vision and control aspects in the design of an HSI framework can lead to severe failures. For instance, failures in estimating human pose, caused by factors such as unbalanced camera vibration or motion blur, can compromise system stability, leading to crashes and potentially posing risks to the op-

[1]Authors are with the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic.

[2]Authors are with the Department of Informatics, Faculty of Sciences, Aristotle University of Thessaloniki, Greece.

⋆Corresponding author: `vit.kratky@fel.cvut.cz`

This work was partially funded by the EU's H2020 AERIAL-CORE grant no. 871479, by the CTU grant no. SGS23/177/OHK3/3T/13, and by the Czech Science Foundation (GAČR) grants no. 23-07517S and 23-06162M.

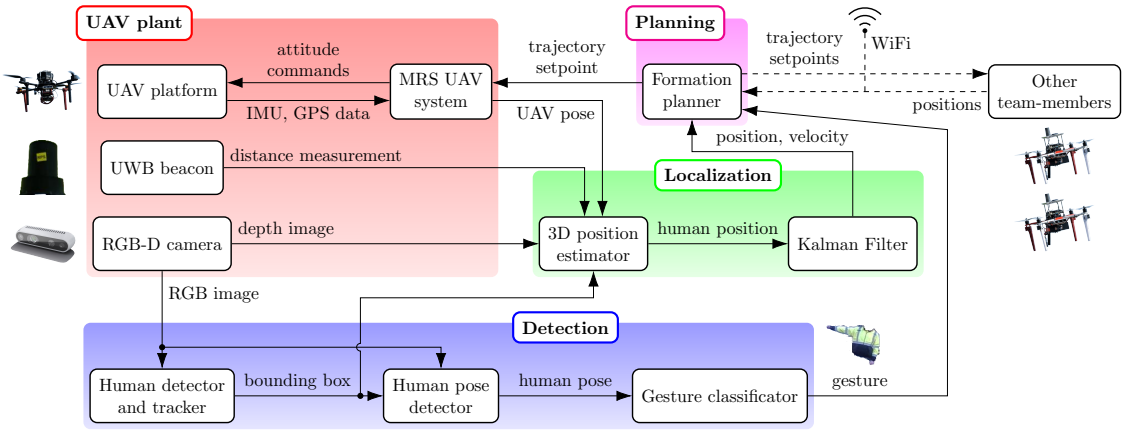

Fig. 2: System architecture overview showing data exchange between blocks using arrows, with highlighted layers.

erator. Importantly, none of the aforementioned studies has addressed the challenge of integrating gesture recognition modules within the UAV formation control scheme to enhance HSI while ensuring rapid responsiveness.

This paper presents a formation control approach for contactless HSI using hand gestures with a team of multirotor UAVs. The approach aims to monitor the status of human workers, particularly those working at heights. In this scenario, one UAV serves as the formation leader, equipped with sensors for human worker detection and gesture recognition. The follower UAVs maintain a predefined formation relative to the worker's position to provide additional views of the scene. Hand gestures enable the human worker to command adaptation of various formation parameters, including relative distance between UAVs, the distance to the worker, relative altitude and direction of the view. Safety measures, such as collision avoidance based on local map information, are also implemented. The proposed approach is validated through field experiments in a mock-up scenario involving three UAVs and a human worker, as illustrated in Figure 1.

III. System Architecture

In this section, we present the system architecture, as depicted in Figure 2, which consists of four layers: *Detection*, *Localization*, *Planning*, and *UAV Plant*. The *Detection* block interacts with the human worker and translates hand gestures into commands for the UAV formation. An RGB-D camera captures images, enabling human detection, tracking, and gesture recognition (Sections III-A and III-B, rexpectively). The *Localization* block combines sensor information from the UAV plant, including the relative distance between the drone and the worker, with data from the *Detection* block and an Ultra Wide Bandwidth (UWB) module. This fusion provides inputs for a Kalman filter, which estimates the 3D position and velocity of the human for the formation controller (Section III-A). The *Planning* block generates feasible trajectories for individual vehicles based on the UAV formation leader's status, the output of the gesture classifier, the human's state, and the status of other UAV team members obtained through a wireless network (Sections III-C and III-D). Finally, the *UAV Plant* receives and executes the trajectories for precise flight [28].

A. Human detection and pose estimation

RGB images from the onboard camera are processed on-the-fly for human detection and tracking, leveraging the authors' prior work on a Convolutional Neural Network (CNN) [29]. A fast deep neural object detector based on Single-Shot multibox Detector (SSD) [30] is employed in combination with a custom LDES-ODDA visual tracker [31]. The output of this pipeline is a predicted bounding box of the tracked human for each input image where the human is visible, as shown in Figure 3. These bounding boxes are then used for gesture recognition and human localization. To maximize accuracy, the detector and the tracker were pretrained on a manually annotated dataset[1] and then finetuned on videos of a human operator wearing safety equipment, captured in diverse outdoor environments and lighting conditions.

The estimation of the tracked human's 3D position relies on various inputs, including the relative direction to the camera, the relative distance to the human, and the pose of the UAV. To determine the *relative direction* of the human from the camera, a pinhole camera model and calibrated camera parameters are employed. Estimating the *distance of the human* involves using three sources: (i) the apparent size in the image based on the bounding box and known physical dimensions of the human, (ii) the depth obtained by taking the median of depth measurements from the stereo camera within the bounding box, and (iii) the distance measurements from the UWB system[2] mounted on both the UAV and the human worker. Additionally, the pose of the UAV is retrieved from onboard sensors. To enhance the accuracy and stability of the estimated 3D position of the human, the Kalman filter is employed. This filter refines the estimated position and also estimates the human's velocity in the world frame, utilizing a constant-velocity first-order point-mass motion model.

[1] https://aiia.csd.auth.gr/open-multidrone-datasets
[2] https://www.terabee.com/shop/mobile-robotics/terabee-robot-positioning-system

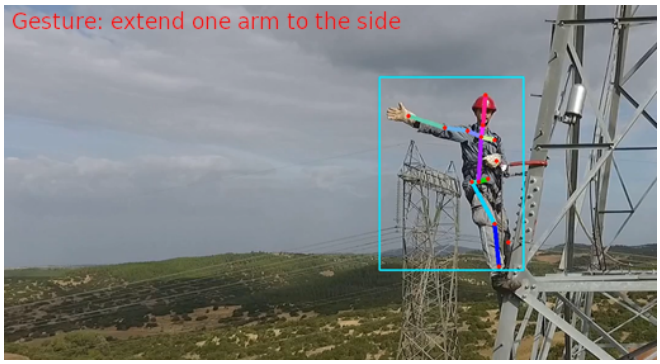

Fig. 3: Example output of the developed gesture recognition pipeline overlaid on the corresponding input video frame.

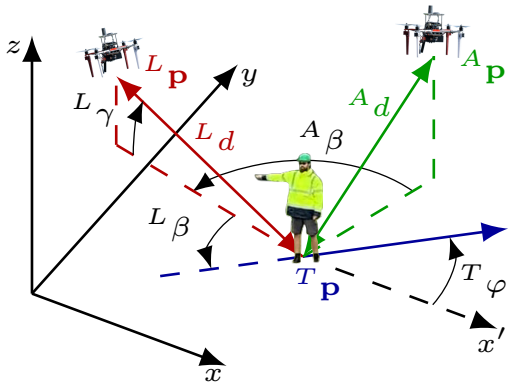

Fig. 4: Illustration of the applied formation scheme for tracking of the human worker while providing the view of the scene from multiple directions and distances.

B. Gesture recognition

Gesture recognition is crucial in enabling visual interaction from the human worker to the UAV formation through gestures. Given a sequence of images captured by the RGB-D camera of the leader UAV and the corresponding bounding boxes of the tracked human (Section III-A), the developed gesture recognition module predicts the type of gesture from a predefined set (e.g., extend one arm to the side) [32], [33]. These predicted gestures are then incorporated into the formation control scheme, as described in Section III-C.

The gesture recognition pipeline consists of two sequential modules. First, 2D human skeletons/poses are extracted using our method described in [34] from the input image, which is cropped by the corresponding bounding box of the tracked human (see Figure 3). The last N outputs of the skeleton extractor are stored in a FIFO buffer. This buffer is subsequently processed by our gesture classifier [35] based on a lightweight Long Short-Term Memory (LSTM) architecture, which outputs the type of the performed gesture.

The pipeline was trained on a large, manually annotated dataset of gestures[3] and finetuned to perform effectively on aerial images, similarly to the human detector described in Section III-A. The parameter N was empirically tuned to $N = 9$ based on the update rate of the camera and the pipeline's performance when running onboard the UAVs.

C. Formation control

The proposed system incorporates a leader-follower formation scheme for formation control, building upon our previous work on aerial filming [6]. In this scheme, depicted in Figure 4, the UAVs maintain a predefined formation relative to the position of the worker detected by the leader, which is shared within the team (see Figure 2). Furthermore, the UAVs ensure that their cameras are directed towards the worker, continuously monitoring the human worker's status.

The state $^i\mathbf{p}$ of the i-th UAV in the formation is described by its position coordinates $^i\mathbf{p} = (^ip_x, ^ip_y, ^ip_z)^\top$ and the orientation of its virtual camera, represented by the heading $^i\varphi$ and pitch $^i\xi$. A label i in the upper left indicates a specific UAV within the team, with $i = L$ referring to the leader UAV, and $i = \{A, B, \dots Z\}/\{L\}$ referring to the follower UAVs. Based on the estimated pose of the target $^T\mathbf{p} = (^Tp_x, ^Tp_y, ^Tp_z, ^T\varphi)^\top$, the desired pose of the leader UAV $^L\mathbf{p} = (^Lp_x, ^Lp_y, ^Lp_z, ^L\varphi, ^L\xi)^\top$ is computed. Given the required observation angles in the horizontal and vertical plane $(^L\beta, ^L\gamma)$ and the required distance to the target (^Ld), the desired state of the leader is given by

$$^L\mathbf{p} = (^T\mathbf{p}^\top, 0)^\top - \begin{pmatrix} ^Ld\cos(^T\varphi - ^L\beta)\cos(^L\gamma) \\ ^Ld\sin(^T\varphi - ^L\beta)\cos(^L\gamma) \\ ^Ld\sin(-^L\gamma) \\ ^L\beta - ^T\varphi \\ ^L\gamma \end{pmatrix}. \quad (1)$$

Similarly, the desired position of the follower UAVs, such as follower A, $^A\mathbf{p} = (^Ap_x, ^Ap_y, ^Ap_z, ^A\varphi, ^A\xi)^\top$, with required observation distance ^Ad and angles $^A\beta$ and $^A\gamma$ defined with respect to leader UAV's camera optical axis, is computed as

$$^A\mathbf{p} = (^T\mathbf{p}^\top, 0)^\top - \begin{pmatrix} ^Ad\cos(^L\varphi - ^A\beta)\cos(^A\gamma - ^L\xi) \\ ^Ad\sin(^L\varphi - ^A\beta)\cos(^A\gamma - ^L\xi) \\ ^Ad\sin(^L\xi - ^A\gamma) \\ ^A\beta - ^L\varphi \\ ^A\gamma - ^L\xi \end{pmatrix}. \quad (2)$$

Note that $^T\varphi$ does not have to match the orientation of worker's body. It can coincide with the direction of estimated motion or can be set to constant value.

The formation scheme, represented by eqs (1) and (2), is applied to every pose on the prediction horizon using the worker's predicted trajectory and the leader's planned trajectory. These trajectories then serve as reference trajectories for a three-stage trajectory generation process depicted in Figure 4. First, collision-free paths along the reference trajectories are generated for each UAV using the map of the environment. Next, safe corridors are computed along these paths using a convex decomposition of free space [6]. Finally, trajectory optimization is performed within each safe corridor to obtain dynamically feasible, collision-free trajectories. The teammates and their planned trajectories are included as obstacles in the map to prevent inter-UAV collisions. This trajectory generation pipeline is executed onboard each UAV in a receding horizon manner, allowing for real-time response

[3] https://aiia.csd.auth.gr/auth-uav-gesture-dataset

to dynamic environment. Detailed information about the trajectory generation process can be found in [6].

The desired observation angles ${}^{A}\beta$ and ${}^{A}\gamma$, and distances ${}^{A}d$ can be set before the mission or adjusted during the flight based on the gestures performed by the worker to achieve the desired view of the scene. The dynamic changes based on the gestures are executed in incremental steps. The trajectory generation algorithm is designed to handle such step changes and produce smooth and feasible trajectories.

D. Gesture processing and formation adaptation

The output of the gesture recognition pipeline undergoes a processing step to improve the reliability of HSI by prevention of undesired shape adaptation due to false positive detections of gestures. During each iteration of the algorithm, only the most recent valid measurements with prediction confidence exceeding a predefined threshold $\Gamma_c \in \mathbb{R}_{>0}$ are considered. The data older than a specified time threshold $t_c \in \mathbb{R}_{>0}$ is filtered out to maintain the relevance of the measurements to current state of the scene. From the remaining set of measurements, the ratio $f_d \in [0,1]$ of the dominant gesture is computed by determining the maximum number of detections associated with a single gesture from the set of gestures, excluding measurements in which no gesture was detected.

If the computed ratio f_d exceeds a predefined threshold $\Pi_d \in [0,1]$, the command for adapting the formation parameter (see Section III-C) corresponding to the dominant gesture is executed. However, a new command can only be executed $t_d \in \mathbb{R}_{>0}$ seconds after the previous call to prevent unwanted repeated updates based on the same set of measurements. This improves the human worker's control of formation parameters updates and prevents the unintentional shape adaptation. The values of Γ_c, Π_d, t_c, and t_d were chosen through multiple real-world experiments, where various initial conditions and sets of gestures were tested.

IV. Experimental results

The effectiveness and validity of the proposed system were demonstrated through field experiments conducted using three UAVs[4] in collaboration with a human worker wearing a reflective safety vest. The following mapping of gestures to the formation parameters was used in the presented experiments: extend arm to side — increase ${}^{L}\beta$, cross arms — decrease ${}^{L}\beta$, raise arm upwards — increase ${}^{L}\gamma$, put palms together — decrease ${}^{L}\gamma$. The increments and decrements of ${}^{L}\beta$ and ${}^{L}\gamma$ were set to 30° and 5°, respectively. The heading of the human worker ${}^{T}\varphi$ was supposed to be constant, and the gestures were filtered using twenty most recent measurements, $t_c = 20\,\text{s}$, $t_d = 5\,\text{s}$, and $\Gamma_d = 0.6$.

During the experiments, the UAV team effectively maintained the desired distance and orientation relative to the worker, ensuring safety and demonstrating the capability of the proposed approach in real-world scenarios. Snapshots of the experiment are shown in Figure 5, providing a visual representation of the UAV team following the worker and adapting the view based on detected gestures. Videos showcasing the real-world demonstration can be accessed at the following link: `http://mrs.felk.cvut.cz/hmri2023gestures`.

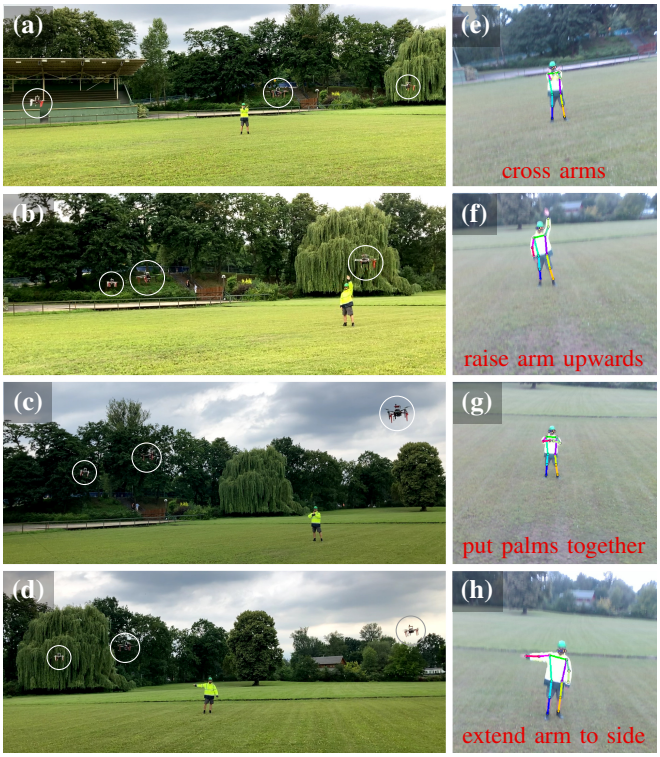

Fig. 5: Sequence of snapshots showing a team of UAVs following a human worker (a)-(d) and adapting the relative view based on the detected gestures (e)-(h).

The conducted experiments highlight the potential of using hand gestures for intuitive control and coordination of autonomous systems. Thus, enabling safe and efficient Human-Swarm Interaction in applications such as safety monitoring and assistance to humans working in difficult-to-access environments.

V. Conclusion

In this paper, we introduced an approach for contactless Human-Swarm Interaction using hand gestures to control a team of UAVs. The proposed approach enables safe and efficient interaction between human workers and autonomous aerial systems, offering benefits in real-world scenarios. The integration of hand gestures as a control modality allows human workers to command and adjust various formation parameters such as relative direction and distance to the worker. The system utilizes robust algorithms for human worker detection and gesture recognition, ensuring accurate and prompt response. Field experiments validated the effectiveness of the approach, demonstrating successful formation control based on detected hand gestures. This work shows the potential for future research in gesture recognition algorithms and tackling scalability challenges for larger swarm formations.

[4]For a detailed description of the hardware platforms see [36], [37].

References

[1] A. Ollero, M. Tognon, A. Suarez *et al.*, "Past, Present, and Future of Aerial Robotic Manipulators," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 626–645, 2022.

[2] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha *et al.*, "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," *IEEE Access*, vol. 7, pp. 48 572–48 634, 2019.

[3] A. Alcantara, J. Capitan, A. Torres-Gonzalez *et al.*, "Autonomous Execution of Cinematographic Shots With Multiple Drones," *IEEE Access*, vol. 8, pp. 201 300–201 316, 2020.

[4] M. Tognon, H. A. T. Chávez, E. Gasparin *et al.*, "A Truly-Redundant Aerial Manipulator System With Application to Push-and-Slide Inspection in Industrial Plants," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1846–1851, 2019.

[5] D. Bonilla Licea, G. Silano, M. Ghogho *et al.*, "Communications-Aware Robotics: Challenges and Opportunities," in *2023 International Conference on Unmanned Aircraft Systems*, 2023, pp. 366–371.

[6] V. Kratky, A. Alcantara, J. Capitan *et al.*, "Autonomous Aerial Filming with Distributed Lighting by a Team of Unmanned Aerial Vehicles," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7580–7587, 2021.

[7] P. Petracek, V. Kratky, M. Petrlik *et al.*, "Large-Scale Exploration of Cave Environments by Unmanned Aerial Vehicles," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7596–7603, 2021.

[8] A. Afifi, M. van Holland, and A. Franchi, "Toward Physical Human-Robot Interaction Control with Aerial Manipulators: Compliance, Redundancy Resolution, and Input Limits," in *2022 International Conference on Robotics and Automation*, 2022, pp. 4855–4861.

[9] S. S. Mansouri, C. Kanellakis, E. Fresk *et al.*, "Cooperative coverage path planning for visual inspection," *Control Engineering Practice*, vol. 74, pp. 118–131, 2018.

[10] G. Loianno, Y. Mulgaonkar, C. Brunner *et al.*, "Autonomous flight and cooperative control for reconstruction using aerial robots powered by smartphones," *The International Journal of Robotics Research*, vol. 37, no. 11, pp. 1341–1358, 2018.

[11] G. Silano, J. Bednar, T. Nascimento *et al.*, "A Multi-Layer Software Architecture for Aerial Cognitive Multi-Robot Systems in Power Line Inspection Tasks," in *2021 International Conference on Unmanned Aircraft Systems*, 2021, pp. 1624–1629.

[12] S. Haddadin, M. Suppa, S. Fuchs *et al.*, "Towards the robotic co-worker," in *Robotics Research*, C. Pradalier, R. Siegwart, and G. Hirzinger, Eds. Springer Berlin Heidelberg, 2011, pp. 261–282.

[13] M. Tognon, R. Alami, and B. Siciliano, "Physical Human-Robot Interaction With a Tethered Aerial Vehicle: Application to a Force-Based Human Guiding Problem," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 723–734, 2021.

[14] F. Benzi, M. Brunner, M. Tognon *et al.*, "Adaptive Tank-based Control for Aerial Physical Interaction with Uncertain Dynamic Environments Using Energy-Task Estimation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9129–9136, 2022.

[15] A. Wojciechowska, J. Frey, S. Sass *et al.*, "Collocated Human-Drone Interaction: Methodology and Approach Strategy," in *2019 14th IEEE International Conference on Human-Robot Interaction*, 2019, pp. 172–181.

[16] A. Ajoudani, A. M. Zanchettin, S. Ivaldi *et al.*, "Progress and prospects of the human-robot collaboration," *Autonomous Robots*, vol. 42, no. 5, pp. 957–975, 2018.

[17] A. Kolling, P. Walker, N. Chakraborty *et al.*, "Human Interaction With Robot Swarms: A Survey," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 9–26, 2016.

[18] A. Dahiya, A. M. Aroyo, K. Dautenhahn *et al.*, "A survey of multi-agent Human–Robot Interaction systems," *Robotics and Autonomous Systems*, vol. 161, no. 104335, pp. 1–18, 2023.

[19] A. Couture-Beil, R. T. Vaughan, and G. Mori, "Selecting and Commanding Individual Robots in a Multi-Robot System," in *2010 Canadian Conference on Computer and Robot Vision*, 2010, pp. 159–166.

[20] J. Nagi, A. Giusti, L. M. Gambardella *et al.*, "Human-swarm interaction using spatial gestures," in *2014 IEEE International Conference on Intelligent Robots and Systems*, 2014, pp. 3834–3841.

[21] V. M. Monajjemi, J. Wawerla, R. Vaughan *et al.*, "HRI in the sky: Creating and commanding teams of UAVs with a vision-mediated gestural interface," in *2013 IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 617–623.

[22] S. Pourmehr, V. M. Monajjemi, R. Vaughan *et al.*, ""You two! Take off!": Creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands," in *2013 IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 137–142.

[23] B. Milligan, G. Mori, and R. Vaughan, "Selecting and commanding groups in a multi-robot vision based system," in *2011 6th IEEE International Conference on Human-Robot Interaction*, 2011, pp. 415–415.

[24] M. Jacquet, M. Kivits, H. Das *et al.*, "Motor-Level N-MPC for Cooperative Active Perception With Multiple Heterogeneous UAVs," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2063–2070, 2022.

[25] P. Petracek, V. Kratky, and M. Saska, "Dronument: System for Reliable Deployment of Micro Aerial Vehicles in Dark Areas of Large Historical Monuments," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2078–2085, 2020.

[26] A. Alcantara, J. Capitan, R. Cunha *et al.*, "Optimal trajectory planning for cinematography with multiple Unmanned Aerial Vehicles," *Robotics and Autonomous Systems*, vol. 140, no. 103778, 2021.

[27] R. Bonatti, W. Wang, C. Ho *et al.*, "Autonomous aerial cinematography in unstructured environments with learned artistic decision-making," *Journal of Field Robotics*, vol. 37, no. 4, pp. 606–641, 2020.

[28] T. Baca, M. Petrlik, M. Vrba *et al.*, "The MRS UAV System: Pushing the Frontiers of Reproducible Research, Real-world Deployment, and Education with Autonomous Unmanned Aerial Vehicles," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 26, pp. 1–28, 2021.

[29] C. Symeonidis, I. Mademlis, I. Pitas *et al.*, "Neural Attention-Driven Non-Maximum Suppression for Person Detection," *IEEE Transactions on Image Processing*, vol. 32, pp. 2454–2467, 2023.

[30] W. Liu, D. Anguelov, D. Erhan *et al.*, "Ssd: Single shot multibox detector," in *In Proceedings of the European Conference on Computer Vision*, 2016, pp. 21–37.

[31] I. Karakostas, V. Mygdalis, A. Tefas *et al.*, "Occlusion detection and drift-avoidance framework for 2D visual object tracking," *Signal Processing: Image Communication*, vol. 90, p. 116011, 2021.

[32] F. Patrona, I. Mademlis, and I. Pitas, "Self-Supervised Convolutional Neural Networks for Fast Gesture Recognition in Human-Robot Interaction," in *2021 10th International Conference on Information and Automation for Sustainability*, 2021, pp. 88–93.

[33] ——, "Gesture Recognition by Self-Supervised Moving Interest Point Completion for CNN-LSTMs," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop*, 2022, pp. 1–5.

[34] C. Papaioannidis, I. Mademlis, and I. Pitas, "Fast CNN-based Single-Person 2D Human Pose Estimation for Autonomous Systems," *In IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 3, pp. 1262–1275, 2023.

[35] C. Papaioannidis, D. Makrygiannis, I. Mademlis *et al.*, "Learning Fast and Robust Gesture Recognition," in *Proceedings of the EURASIP European Conference on Signal Processing*, 2021, pp. 761–765.

[36] D. Hert, T. Baca, P. Petracek *et al.*, "MRS Modular UAV Hardware Platforms for Supporting Research in Real-World Outdoor and Indoor Environments," in *2022 International Conference on Unmanned Aircraft Systems*, 2022, pp. 1264–1273.

[37] D. Hert, T. Baca, P. Petracek *et al.*, "MRS Drone: A Modular Platform for Real-World Deployment of Aerial Multi-Robot Systems," *Journal of Intelligent & Robotic Systems*, vol. 108, no. 64, pp. 1–34, 2023.