Abstract—Power, size and weight are major constraining factors for UAVs, which have caused researchers to investigate alternative information representations, such as semantics, to reduce the computational load onboard UAVs. However, the connection between spatial and semantic information has not been fully explored and the literature on including semantic information for UAV navigation is limited. This paper provides a new taxonomy on spatial and semantic relations through consideration of a selected subset of the literature. Concepts of semantic information are related with spatial maps and two hypotheses for improving algorithms based on the literature are generated. The algorithms are tested in numerical simulations on the task of semantic mapping and the effects of influential parameters are examined. Results indicate that a defined hierarchical structure of semantic information enables improved map reproduction. The proposed algorithm exhibits the potential to improve the use of semantic information in UAV deployment scenarios, such as agriculture, search and rescue and industrial inspections.

TABLE OF CONTENTS
1. INTRODUCTION ................................... 1
2. LITERATURE TAXONOMY ........................ 2
3. SIMULATIONS .................................... 3
4. CONCLUSIONS .................................... 7
ACKNOWLEDGMENTS ................................. 8
REFERENCES ........................................ 8
BIOGRAPHY ......................................... 11

1. INTRODUCTION

Limitations in power, size and weight in UAVs have resulted in researchers exploring alternative approaches of representing information to reduce the computational load onboard UAVs. Contextual information is often used for navigation, with semantics at the frontier of contemporary robotics research showing promising results. Kostavelis and Gasteratos define semantics as “... related to the study between signs and the things to which they refer, that is their meaning,” [1]. Semantic information has been included into the passive part of SLAM [2], [3], as well as into active path planning [4], [5]. However, in terms of semantic signal source, multiple levels of signals are possible, such as individual objects, regions or areas [1]. These cause new problems for defining successful navigation, as highlighted by Anderson et al. [6]. Furthermore, recent advances in Computer Vision have produced a wide range of different sensors signals to be annotated with semantic information, such as pixel-level segmentation, object detection or scene classification [7], which operate on different spatial levels. The influence of these sensor signals on the level of names on conventional robotics algorithms, such as filters and path planners, is an active field of research [1], [2], [6]. UAVs are impacted by unique viewing angles, as well as risk-averse requirements and hard constraints on computational resources [8], [9], [10], which add additional complexity to the problem of semantic based navigation. Figure 1 illustrates the context where the UAV has a field of view and semantic classes are estimated from a top-down view and assigned to a map of grid cells. The classes for this example are house, vehicle, grass, road and tree.

Two key contributions are presented, the first of which leads to the second. Firstly, a review of a selected subset of the literature in this area is conducted to create a taxonomy that enables the key factors in successful utilization of semantic information. Using these new insights, algorithms are proposed and evaluated in functionally-relevant numerical simulations according to a generic autonomous algorithm development framework [11]. Initial results demonstrate the potential of the approach to improve semantic mapping performance.

This paper is split into two parts: the first part is concerned
with a taxonomy of a selected subset of literature on semantic navigation. The second part contains the proposal of algorithms based on conclusions drawn from the literature, which are evaluated in numerical simulations. Each of these sections contain their own structure of Methodology, Results and Discussion, with a final Conclusion section summarizing the results and providing an outlook. This research is directed towards enhancing the performance of a module which is tasked with extracting and evaluating the semantic information during UAV path planning in a navigational framework [12].

## 2. LITERATURE TAXONOMY

In this paper we evaluate a selected subset of literature, which combines semantic signals with spatial information in robotics to demonstrate the state of the research. The literature is categorised with respect to different dimensions.

### Methodology

Twenty-six papers were selected for this research. The papers are sorted according to the following three dimensions and the methods employed in those papers are extracted:

1. Robot Type
2. Spatial Map
3. Semantic Map

Significant research is briefly summarized along with the rough timeframes of the projects to contextualize the progress in the field. The main contributions are extracted and the form of inclusion of semantic information with spatial information is highlighted.

### Results

The taxonomy of the selected literature illustrates the methods employed in the literature with respect to the spatial, semantic and robot type-dimensions.

Figure 3 shows the results of generating a taxonomic representation of the included literature. The literature extends to a selection from 1995 to 2020 and has been classified according to the spatial and semantic maps that were used, as well as the robot type employed in the research. It should be noted that in most cases image space was used to extract semantic information, however, the map retention is considered as the defining factor for the spatial map. The majority of research considered in this work is employing a hybrid approach on either the semantic level or the spatial level. A smaller portion considers non-hybrid information and is placed in the corners of the figure.

### Generic Research—

Five research efforts considered as “generic” from as early as 1995 have been included in the taxonomy of this paper.

Chown and colleagues for example [13] researched the "what" and "where" subsystems of the human cognition and their learning over time. They captured the influence of landmark recognition, recreation of consistent Euclidean maps, as well as the importance of gateways, representing transitions between regions. Suenderhauf et al. [14] associated point-clouds in SLAM with labels propagated through a modern neural network. Yang et al. [15] used scene priors derived from vector representations of class names to enhance a reinforcement learning module which relies on visual cues in indoor environments. Crespo et al. [16] generalised the implementation of a knowledgebase-system developed by Tenorth et al. [17] to a database implementation which improved the speed significantly. In 2018, Wu et al. [18] used reinforcement learning to train an agent to recognise a room and infer whether one room type is directly or indirectly accessible from another room using a Bernoulli distribution.

### Ground Robots—

Eight research efforts from a timespan between 2000 and 2020 have been considered in this taxonomy. Contemporary image processing techniques have accelerated semantic spatial research through significant improvements in reliability and robustness of classification and localisation in image space [7].

Kuipers [19] connected places, defined as 0D “points”, through 1D "paths" with 2D "regions" in topological maps. The proposed maps had a sensory and control level, a causal level, a topological level and a metrical level, all of which were connected hierarchically. Together with his colleagues, Kuipers [20] also conducted experiments where local metric room representations were linked through topological constraints. Galindo et al. [21] connected a spatial hierarchical representation consisting of areas and objects with a terminological box and connected between these two in a deterministic manner to enable inference and path planning. Borkowski et al. [22] assigned a hierarchical place taxonomy based on house representations from architecture to a metric map and performed path-planning within this map while considering semantic constraints. Tenorth [17] built a topological representation of abstract objects and locations and associates these in a hierarchical fashion to enhance task planning by incorporating the relations of objects. Krishnan and Krishna [23] constructed a hierarchical semantical-topological and explored the semantic nodes before proceeding to the next. Zhang et al. [3] used a hierarchical topic model to improve the performance of a SLAM system by actively integrating the object association problem into a hierarchical Dirichlet process.

Chaplot et al. [24] won the 2020 CVPR challenge to navigate to an object goal by learning a semantic-metric map from visual cues and training a policy to generate frontier-based goals. Their research also highlighted that other Reinforcement Learning policies were outperformed by frontier-based methods and were unable to generalize to the real-world.

### UAVs—

Thirteen UAV-based research efforts from 2013 onward have been included in this taxonomy, which present a variety of approaches of including semantic information.
A substantial amount of literature is concerned with passive acquisition of semantic information:

Le Saux and Sanfourche [25], Sheppard and Rahmoomifar [26], Christie et al. [27] and Kyrkou et al. [28], classified top-down images taken from a UAV. Cavaliere et al. [29], [30] associated “tracks” coming from images with “places” from geospatial information to describe relations between them in a relational manner, as detailed in [31]. The individual items were structured hierarchically to distinguish between different type of tracks, such as vehicles and humans on the first level and cars, motorcycles and trucks on the second level. Drouilly et al. [32] developed a metric to evaluate the semantic quality of a path through an environment depending on the distribution of objects and the observation quality.

Semantic information has also been employed in active navigation. Maturana et al. [33] annotated a 2.5D map with detection of cars and developed a system to approach the cars using UAVs. Maravall et al. [34] used the image entropy to navigate through a topological map to semantically salient places. Alirezaie [5] developed a RRT implementation that discards semantically inadmissible points through geospatial information associated with the waypoints and significantly improved planning time for longer distances. Dang et al. [35] employed an object detector on top of a voxel grid to detect humans and bicycles. Koch et al. [4] respected semantic constraints during the path planning step for 3D reconstruction using UAVs while maintaining the quality of the reconstruction. Toudeshki et al. [36] combined a visual teach-and-repeat approach with an object detector to funnel the UAV to follow a path.

Summary—Fusing semantic information with spatial information for navigation, mapping or task-planning has been an active field of research for the past decades. A majority of research fuses information into hybrid models, either on the semantic or on the spatial axis. This is demonstrated by the centralised cluster as shown in figure 3. UAV research is more spread out than other approaches in this figure.

The centralised clustering indicates that research has benefited from hybrid approaches, which have not yet been translated to UAVs due to the hurdle of additional complexity. A significant portion of the included literature incorporated a hierarchical structure, most commonly in the form of an associative connection between “areas” – or “rooms” – and “objects” [21], [3], [24], [17], [22], [19], [37], [29]. Based on these insights a hybrid algorithm, which is to be tested in a simulation scenario with reduced complexity, is proposed.

3. Simulations

The literature taxonomy uncovered that a substantial part of semantics research in robotics benefits from hybrid approaches along the spatial and semantic axes, denoted by the centralised cluster in figure 3. UAVs are constrained by additional layers of complexity in translation of conventional
Figure 4: Examples of the simulations abstracted from possible aerial images, see figure 1. Colors are intended to represent different semantic classes according to table 1, see figure 1 for context. a) shows the first simulation environment $s = 1$, b) the second environment $s = 2$, c) a transposed configuration, $t = 1$, e) randomised configuration $r = 1$ and f) a test configuration $c = 1$ as defined in table 2. d) shows an example path, with the $\times$ representing the UAV, the green cells the previously seen cells, the red cells the currently seen and already known cells and the blue cells the cells predicted at $t-1$.

The green cells in the figure indicate the cells that have already been observed, the red cells the current $f$ of the UAV, which is marked by the black $\times$. The blue cells denote the cells where the prior $p(x)$ has been predicted at $t-1$, based on the posterior of the red cells.

Four algorithms are compared in parallel with the same sampled observations. In all cases, only the prior $p(x)$ for each cell is predicted and all subsequent observations are incorporated according to equation 1. In the baseline case (algorithm 1), the prior for each cell $p(x)$ is uniform over all classes $K$, which is denoted as a Naive Bayes Filter. In the other three cases, spatial correlation is incorporated into the prior. The prediction for the cells is made by extracting the next waypoint $t+1$ and evaluating all cells that will be visible within the field-of-view $f$ at that timestep. In the second algorithm (algorithm 2), yet unobserved cells receive a prior which is a linear combination of the uniform prior and the posterior of the already observed cells. The posterior probability is weighted by a mixing parameter $\alpha \in [0, 1]$, as well as the ratio of number of observed cells divided by total cells – defined through the overlap $o$ and $f$ – and used as a prior for the yet unobserved cells.

The other two algorithms are based on assumptions drawn from the results of the literature taxonomy as detailed in section 2. Incorporating information in a hierarchical context has proven successful in related semantic fields, such as speech processing [39] and topic modelling [40]. These approaches have been successfully adapted to related domains, such as image annotation [41], [42] or SLAM [3]. Some algorithms have incorporated the hierarchical semantic information successfully for the purpose of navigation [4], [24], [18], [5].
Two algorithms that recalculate the prior based on hierarchical information are proposed: the prior is calculated according to an assumed distribution of "areas" – u – groupings of cells. These groupings require the definition of \( p(x \mid u) \) – the distribution of "objects" in a given "area" – and \( p(u) \), as well as the number of groupings, which is usually smaller than \( K \), analogous to topic modelling [40]. In the third algorithm (algorithm 3), every map cell is assigned a fixed prior of \( p(x_{t=0}) = p(x \mid u)p(u) \). In the fourth algorithm (algorithm 4) \( p(u) \) is updated dynamically by using the observations of each cell within \( f \), with a fixed \( p(x \mid u) \), which leads to a new dynamically assigned \( p(x) \).

Three variations of \( p(u) \) and \( p(x \mid u) \) – denoted by the parameter \( Ptu \) – are considered. The first is a manual version, where the mixture of "objects" in "areas" \( p(x \mid u) \) is estimated by the user and the distribution of "areas" \( p(u) \) in the map as well. In the second and third version, \( p(x \mid u)\) and \( p(u) \) are derived through Monte-Carlo simulations described in section 3. \( K = 5 \) and 3 are used for the simulations, because the small number allows for interpretations, visual simplicity and numerical stability when defining and evaluating simulation cases.

Robotics research is often constrained by the overlap of classes available in detection and simulation environments [24], such that 5 is a reasonable approximation for numerical research. The chosen classes are distributed according to different scales, with one class being dominant in the map, occupying over 70 % of the cells in the majority of cases, while another class occupies less that 1 %. Furthermore, the classes occupy different connected regions of the map, ranging from 1 isolated cell to multiples of the field-of-view, which allows the investigation how different scales influence the performance.

The simulation cases can be transposed – denoted by \( t \) in figure 4 c)– to evaluate the influence of the path planning direction on the predictive performance. The second simulation environment has further Boolean parameters, that randomize the placement of objects in quadrants \( r \), as well as a parameter that mixes two quadrants \( c \) to illustrate visual diversity and reduce uniformity of areas.

The simulation cases are displayed in figure 4 and table 2 shows the parameters that have been varied. The mixing parameter \( \alpha \) is evaluated in a preliminary set of experiments, where \( \alpha \in [0, 0.2, \ldots, 1] \) and therefore not included in table 2.

Figure 4 illustrates the basic variations for the simulation environments. 4 a) is the first environment \( s = 1 \), which is also used for the Monte-Carlo simulations – detailed below – and excluded from the evaluation. 4 b) shows the second environment \( s = 2 \), and 4 c) a transposed environment \( t = 1 \). 4 e) shows a case where the placement of objects within their quadrants is randomised \( r = 1 \). 4 f) shows a test environment \( c = 1 \) with partial transpose, which convolutes the definition of an "area" as a section identifiable by its composition of "objects". The colours are indicators of types of classes and can be interpreted as shown in table 1, however, are not limited to the named classes. 4 d) shows an example path, with the green colour indicating the already observed fields, red the already seen fields at \( t = 1 \) and blue the previously unseen fields, which are predicted. All combinations are used for the simulations.

Monte-Carlo simulations—1000 simulations with fixed \( o = 0.5, f = 2 \) and

\[
p(z \mid x) = \begin{cases} 
0.8, & \text{for } k = x \\
\frac{0.1}{1-k}, & \text{otherwise}
\end{cases}
\]

are run on the first environment. \( p(u) \) and \( p(x \mid u) \) are sampled from a simplex and the performance of each algorithm as defined in section 3 and \( p(u) \) and \( p(x \mid u) \) are saved if:

1. The log-likelihood is smaller than the current smallest
2. The difference between the predicted algorithm and the dynamic algorithm is bigger than the current biggest

The final best results are used as parameters \( p(u) \) and \( p(x \mid u) \) for additional simulations as defined in section 3.

Evaluation—The reproduction quality of the posterior map for each individual state was evaluated with the sum over the log-likelihoods for each cell \( m_{i,j} \), defined by:

\[
H = - \sum_{i,j,k} p_{i,j,k} \ln(q_{i,j,k})
\]

(2)

is calculated, where \( p \) represents the true state and \( q \) the estimated map state.

For the comparison of different algorithms running with the same parameters, the log-likelihoods for each individual case are subtracted from another and the count of the positive and negative samples indicate the relative quality of the reproduction:

\[
\frac{pos}{pos + neg} \in [0, 1]
\]

(3)

If the value is below 50\%, the algorithm performs worse than its compared baseline. This metric is chosen as a robust indicator, which is not affected by large variances for each simulation run caused by the different parameters. Only the second simulation environment is included in the evaluation,
since the first simulation environment was used in the MonteCarlo simulation to generate sample probability vectors, as described above.

Results

Numerical simulations were conducted to test the performance of the algorithms under different conditions. Figure 5 displays the results of the numerical simulations evaluated on the second simulation environment, \( s = 2 \). Preliminary experiments showed that the second algorithm, which predicts yet unobserved cells with a mixture of \( \alpha = 0.5 \) always outperformed the weakest baseline and largely outperformed the second, hierarchical baseline. Therefore, results are presented which compare the hierarchical dynamic prediction of \( p(x) \) through \( p(u) \) and \( p(x | u) \) with the prediction through the mixture component of \( \alpha = 0.5 \). A total of 1975 simulation cases are included in this evaluation. The height of the bar indicates the percentage of outperformance, as indicated by equation 3. A value below the threshold of 50% indicates a worse performance. The bars for \( \text{Dim} \) and \( \text{Fov} \) – \( \text{f} \) in table 2 – indicate that neither the dimensions of the map \( M \), nor an extended field of view \( f \) lead to an improvement of one algorithm over the other. \( \text{Transp} \) and \( \text{Test} – t \) and \( c \) in table 2 respectively – also do not significantly influence the result.

The decreasing bars for the accuracy-parameter in figure 5 indicate that the influence of the prior diminishes when the detection accuracy improves, which is aligned with the intuition of Bayesian filtering, as the detections converge to the true state. The increasing performance with increasing overlap \( o \) – which is running in parallel for both axes – show that the additional observations benefit the hierarchical algorithm more than a linear mixing. The higher value for the \( \text{Rand} \) parameter – \( r \) in table 2 – show that even in a randomized setting, the hierarchical prediction can lead to increased performance. The low value for \( 
\text{Ptu} \) case 0 shows that a manually defined detection parameter performs significantly worse, however, the high values for 1 and 2 indicate that with the right parameters for \( p(u) \) and \( p(x | u) \), the performance can increase significantly.

Figure 6 shows two reproductions of example cases, with the entropy value of the reproduction calculated by equation 2 in the title. Lower values indicate more accurate reproductions. Both cases show the second simulation environment used for testing, \( s = 2 \), with an overlap \( o = 0.5 \), \( f = 1 \) and \( t = 0 \) as defined in table 2. As can be seen, both algorithms perform much better than the baseline, which is indicated by the noise in the reproduction. The first row shows an example case where the proposed algorithm was outperformed by algorithm 2, albeit by a small margin. This case represents a test configuration of \( c = 0, r = 0 \) accuracy of 0.8 and \( 
\text{Ptu} = 0 \). In the second row the parameters are: \( c = 1, r = 1 \), accuracy 0.7 and \( 
\text{Ptu} = 2 \). Despite performing better than the other algorithms, the reproduction is much worse than with a higher detector accuracy which points to the significant influence of this parameter. This is shown by the large gap between row I and II for the Naive Bayes classifier, which assumes independence between individual cells and is therefore not influenced by spatial variations \( c \) and \( r \). However, in row II the log-likelihood decreases with a change in algorithm, despite the randomness and convolution throughout the map, which indicates that the proposed algorithm increases the performance, if the right spatial parameters can be retrieved. Manually defining these, despite best guesses, does not yield
Figure 6: Two reproduction examples for different parameters as presented in table 2. The titles indicate the algorithm, as well as the total log-likelihood as defined by equation 2. Both cases show results for the second simulation environment \( s = 2 \) with an overlap \( o = 0.5 \), \( f = 1 \) and \( t = 0 \). Row I shows the results for \( Ptu = 0 \) accuracy 0.8 and \( c = 0 \). Row II for \( Ptu = 2 \), accuracy 0.7 and \( c = 1 \). Columns a) show the ground truths, b) the Naive Bayes filter, c) Algorithm 2 and d) the Proposed Algorithm 4.

4. CONCLUSIONS

This work has categorized twenty-six papers with respect to spatial map, semantic map and robot type for the purpose of gathering information on the benefits and challenges in image space, topological and metric representations. Subsequently, algorithms have been proposed that are to be used in the development pipeline shown in figure 2. The algorithms have been tested in numerical simulations before proceeding with further development steps.

**Literature Taxonomy**

The taxonomy of twenty-six papers, even though not exhaustive, highlights significant factors that influence the synergy between semantic and spatial information. Not targeting robotics and therefore not included in the taxonomy is the relevant work by Harnad on symbolic systems [43], which discusses the association of sensory inputs with symbolic systems and intricate representations. Kostavelis and Gasteratos have completed an extensive literature review of semantic mapping for mobile robotics [1] and Landsiedel et al. [31] have reviewed spatial relations in robotics. The taxonomy presented in this paper highlights the dimensions of semantic and spatial information with respect to the robot type and possible trends to drive the development of an algorithm.

The clustering of UAV-based semantic research in the corners of figure 3 indicates that mixing hierarchical information for UAVs is not an easy task. While enormous datasets and challenges along with unified metrics have driven computer vision research [7], [44] and are starting to show the benefits in navigation [6], [24], UAVs are facing additional challenges. On the one hand the viewpoint of UAVs compared to generalised image datasets is unique, with oblique and top-down viewing points decreasing the accuracy of detectors [45]. On the other hand, UAVs face additional obstacles with safety considerations, as well as reduced computational capabilities [10], [8]. Furthermore, increasingly complex architectures and differing quality requirements can cause difficulties in deploying algorithms [46], [47]. Semantic research generally benefits from unified metrics [6], simulation environments and challenges [24], [15]. The driving force of standardised environments and challenges [44] can also be seen in the UAV community, as exhibited by the development of autonomous drone racing algorithms [48].

Chown et al. [13] emphasize the importance of "gateways" as transitions between regions, which arise naturally in man-made houses and have been used in many research efforts for ground-based vehicles [20], [18], [23], however, may be difficult to define for UAVs. Where an area ends and another begins may be ambiguous for large outdoor environments, which are usually covered by UAVs [49], [38], whereas rooms often provide a natural separation in indoor environments. Furthermore, definitions of semantic area, region, or scene, if representing a spatial extent, could not be identified from the literature and the borders may be fluid between all of these definitions depending on the task and research field.
Simulations

The results of the simulations indicate that incorporating spatial correlation can benefit the assignment of class names to grid cells, which is aligned with robotic occupancy grid mapping approaches, as well as the intuition behind convolutions and filter kernels. The inclusion of explicit hierarchical information into a mapping pipeline can be beneficial, but is largely dependent on the robustness of the model, much like topic modelling approaches [41], [40].

In this research, important hardware parameters such as flying velocity and battery charge [10] have been omitted for simplicity. Furthermore, the simulation environment is a reduced to a manually-defined version, because no standardised simulation environment for semantic mapping for UAVs could be identified and therefore exhibits a natural bias. The simplicity has been attempted to be counteracted through the consideration of parameters $c$, $t$ and $r$ to increase the diversity of test cases. The reliance on detection parameters is demonstrated through the significant drop in performance for $Ptu$ as seen in figure 5 and the convergence of the performance with increasing accuracy. With perfect detectors the filtering approach would not be necessary, however, despite increasing performance [7] the quality of conventional detections is decreased for UAVs due to unique viewpoints [45] and computational constraints [50]. The hierarchical approach proposed in this research depends on the parameters $p(u)$ and $p(x|u)$, as clearly indicated by figure 5, which could be learned from data with an appropriate generative model [40], [42].

**Future Work**

Learning the parameters of the hierarchical model would require a formalised definition of the generative probabilistic model. Semantics, as the study of signs and the things to which they refer [1], has profound roots in language. Computer scientists working with language have put large efforts into attempting to model spoken and written language. Spoken language has employed hierarchical models of language based on the Markov assumption to separate spoken words [39], [51], which have been translated to alternative purposes, such as motion prediction [52] and motion recognition [53], [54]. Researchers in the field of Natural Language Processing have put significant efforts into classifying words into topics and documents (which are considered mixtures of topics) for classification and proposition [40], [55], which has been adapted for the task of scene annotation in computer vision [56], [41] and on satellite images [42]. Learning the parameters from data analogous to these models would enable the model to be extended further, such as through the incorporation of location uncertainty.

Further research will focus on progressing the model through an appropriate generative model and an appropriate simulation environment for semantic mapping for UAVs. Subsequent tests as defined in figure 2 will enable an embedding into higher-level frameworks for semantic navigation [12], [57] and finally deployment on a UAV [8].

**ACKNOWLEDGMENTS**

This research was conducted by the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016) and supported by the QUT Centre for Robotics. The authors acknowledge continued support from the Queensland University of Technology (QUT) through the Centre for Robotics. N. M. acknowledges M. Xu for his technical support during the development of this work and QUT for the financial support.

**REFERENCES**

[1] I. Kostavelis and A. Gasteratos, “Semantic mapping for mobile robotics tasks: A survey,” *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889014003030

[2] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[3] J. Zhang, M. Gui, Q. Wang, R. Liu, J. Xu, and S. Chen, “Hierarchical Topic Model Based Object Association for Semantic SLAM,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 11, pp. 3052–3062, Nov. 2019.

[4] T. Koch, M. Körner, and F. Fraundorfer, “Automatic and Semantically-Aware 3D UAV Flight Planning for Image-Based 3D Reconstruction,” *Remote Sensing*, vol. 11, no. 13, p. 1550, Jan. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/13/1550

[5] M. Alirezaie, A. Kiselev, F. Klügl, M. Längkvist, and A. Loutfi, “Exploiting context and semantics for UAV Path-finding in an urban setting,” vol. 1935, 2017, pp. 11–20.

[6] P. Anderson, A. Chang, D. S. Chaplot, A. Dosovitskiy, S. Gupta, V. Koltun, J. Kosecka, J. Malik, R. Mottaghi, M. Savva, and A. R. Zamir, “On Evaluation of Embodied Navigation Agents,” *arXiv:1807.06757 [cs]*, Jul. 2018, arXiv: 1807.06757. [Online]. Available: http://arxiv.org/abs/1807.06757

[7] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Eesjn, A. A. Š. Awwal, and V. K. Asari, “The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches,” *arXiv:1803.01164 [cs]*, Sep. 2018, arXiv: 1803.01164. [Online]. Available: http://arxiv.org/abs/1803.01164

[8] N. Mandel, M. Milford, and F. Gonzalez, “A Method for Evaluating and Selecting Suitable Hardware for Deployment of Embedded System on UAVs,” *Sensors*, vol. 20, no. 16, p. 4420, Jan. 2020, number: 16 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/20/16/4420

[9] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston, “Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation,” *Sensors*, vol. 16, no. 1, p. 97, Jan. 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/1/97

[10] B. Boroujerdian, H. Genc, S. Krishnan, W. Cui, A. Faust, and V. Reddi, “MAVBench: Micro Aerial Vehicle Benchmarking,” in *2018 51st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Oct. 2018, pp. 894–907.
I. Malatova, G. Lewis, B. Schmerl, P. Lago, E. Cervera, “Try to Start It! The Challenge of Reusing B. Richard Webster, S. E. Anthony, and W. J. Scheirer, A. G. Toudeshki, F. Shamshirdar, and R. Vaughan, “R o- D. Cavaliere, A. Saggese, S. Senatore, M. Vento, and V. Loia, “Empowering UAV scene perception by semantic spatio-temporal features,” in 2018 IEEE International Conference on Environmental Engineering (EE), Mar. 2018, pp. 1–6. S. J. Shetty, R. Ravichandran, L. A. Tony, N. S. Abhinay, K. Das, and D. Ghose, “Implementation of Survivor Detection Strategies Using Drones,” arXiv:2003.12559 [cs, eess], Apr. 2020, arXiv: 2003.12559. [Online]. Available: http://arxiv.org/abs/2003.12559 S. Fine, Y. Singer, and N. Tishby, “The Hierarchical Hidden Markov Model: Analysis and Applications,” Machine Learning, vol. 32, no. 1, pp. 41–62, Jul. 1998. [Online]. Available: https://doi.org/10.1023/A:1007469218079 D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” The Journal of Machine Learning Research, vol. 3, no. null, pp. 993–1022, Mar. 2003. [Online]. Available: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf L. Fei-Fei and P. Perona, “A Bayesian hierarchical model for learning natural scene categories,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 2, Jun. 2005, pp. 524–531 vol. 2, iSSN: 1063-6919. M. Lienou, H. Maître, and M. Datcu, “Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation,” IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 1, pp. 28–32, Jan. 2010, conference Name: IEEE Geoscience and Remote Sensing Letters. S. Harnad, “The symbol grounding problem,” Physica D: Nonlinear Phenomena, vol. 42, no. 1, pp. 335–346, Jun. 1990. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167278990900876 P. Corke, F. Dayoub, D. Hall, J. Skinner, and N. Sünderhauf, “What can robotics research learn from computer vision research?” arXiv:2001.02366 [cs], Jan. 2020, arXiv: 2001.02366. [Online]. Available: http://arxiv.org/abs/2001.02366 B. Richard Webster, S. E. Anthony, and W. J. Scheirer, “PsyPhy: A Psychophysics Driven Evaluation Framework for Visual Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 9, pp. 2280–2286, Sep. 2019. E. Cervera, “Try to Start It! The Challenge of Reusing Code in Robotics Research,” IEEE Robotics and Automation Letters, vol. 4, no. 1, pp. 49–56, Jan. 2019. I. Malatova, G. Lewis, B. Schmerl, P. Lago, and D. Garlan, “How do you Architect your Robots? State of the Practice and Guidelines for ROS-based Systems,” in ICSE2020 - Software Engineering in Practice, vol. 42. Seoul, South Korea: ACM, 2020. [Online]. Available: https://conf.researchr.org/details/icse-2020/icse-2020- Software-Engineering-in-Practice/9/How-do-you-Architect-your-Robots-State-of-the-Practice-and-Guidelines-for-ROS-based- H. Moon, J. Martinez-Carranza, T. Cieslewski, M. Faessler, D. Falanga, A. Simovic, D. Searamunza, S. Li, M. Ozo, C. De Wagger, G. de Croon, S. Hwang, S. Jung, H. Shim, H. Kim, M. Park, T.-C. Au, and S. J. Kim, “Challenges and implemented technologies used in autonomous drone racing,” Intelligent Service Robotics, vol. 12, no. 2, pp. 137–148, Apr. 2019. [Online]. Available: https://doi.org/10.1007/s11370-018-00271-6 F. Vanegas, D. Bratanov, K. Powell, J. Weiss, and F. Gonzalez, “A Novel Methodology for Improving Plant Pest Surveillance in Vineyards and Crops Using UAV-Based Hyperspectral and Spatial Data,” Sensors, vol. 18, no. 1, p. 260, Jan. 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/1/260 S. Krishnan, Z. Wan, K. Bhardwaj, P. Whatmough, A. Faust, G.-Y. Wei, D. Brooks, and V. J. Reddi, “The Sky Is Not the Limit: A Visual Performance Model for Cyber-Physical Co-Design in Autonomous Machines,” IEEE Computer Architecture Letters, vol. 19, no. 1, pp. 38–42, Jan. 2020, conference Name: IEEE Computer Architecture Letters. K. Murphy, “Dynamic Bayesian Networks: Representation, Inference and Learning,” PhD Thesis, University of California, Berkeley, Berkeley, CA, 2002. [Online]. Available: https://www.cs.ubc.ca/~murphyk/Thesis/thesis.pdf Q. Li and H. C. Lau, “A Layered Hidden Markov Model for Predicting Human Trajectories in a Multi-floor Building,” in 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 2, Dec. 2015, pp. 344–351. N. Oliver, E. Horvitz, and A. Garg, “Layered representations for human activity recognition,” in Proceedings, Fourth IEEE International Conference on Multimodal Interfaces, Oct. 2002, pp. 3–8. D. Aarno and D. Kragic, “Layered HMM for Motion Intention Recognition,” in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 2006, pp. 5130–5135, iSSN: 2153-0866. T. Hofmann, “Unsupervised Learning by Probabilistic Latent Semantic Analysis,” Machine Learning, vol. 42, no. 1, pp. 177–196, Jan. 2001. [Online]. Available: https://doi.org/10.1023/A:1007617005950 A. Bosch, A. Zisserman, and X. Muñoz, “Scene Classification Via pLSA,” in Computer Vision – ECCV 2006, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer, 2006, pp. 517–530. F. Vanegas and F. Gonzalez, “Uncertainty based online planning for UAV target finding in cluttered and GPS-denied environments,” in 2016 IEEE Aerospace Conference, Mar. 2016, pp. 1–9, iSSN: null.
**Biography**

**Nicolas Mandel** received the BSc in Sports and Technology from the Otto-von-Guericke University Magdeburg and the MSc in Biomedical Engineering Sciences from the University of Applied Sciences Technikum Vienna after writing his thesis about mechanical modelling of cardiovascular treatments at the University of Cape Town. He is currently pursuing his PhD with the ARC Centre of Excellence for Robotic Vision at Queensland University of Technology, Brisbane, focussing on employing semantics for UAV navigation.

**Michael Milford** received the Ph.D. in Electrical Engineering and the Bachelor of Mechanical and Space Engineering from the University of Queensland (UQ), Brisbane, Australia. He is currently Professor of Robotics, Chief Investigator for the Australian Research Council Centre of Excellence for Robotic Vision and Deputy Director of the QUT Centre for Robotics at the Queensland University of Technology, Brisbane, Australia. His research models the neural mechanisms in the brain underlying tasks like navigation and perception to develop new technologies in challenging application domains such as all-weather, anytime positioning for autonomous vehicles.

**Felipe Gonzalez** is an Associate Professor at Queensland University of Technology and a CI in the Centre for Robotics where he leads the QUT Airborne Systems Lab. He has a BEng(Mech) and a PhD from the University of Sydney. His research explores bioinspired optimization, uncertainty based UAV path planning and UAV based remote sensing. He currently leads two ARC Discovery Projects. His research focuses on developing UAVs with high levels of cognition to assist in understanding, exploring and protecting natural and physical environments. Felipe is a Chartered Professional Engineer and member of professional organizations including the RAeS, IEEE and AIAA.