Abstract—In this article, we propose a hierarchical data collection scheme, toward the realization of unmanned aerial vehicle (UAV)-aided industrial wireless sensor networks. The particular application is that of agricultural monitoring. For that, we propose the use of hybrid compressed sampling through exact and greedy approaches. With the exact approach—to model the energy-optimal formulation—an improved linear programming formulation of the minimum cost flow problem was utilized. The greedy approach is based on a proposed balance factor parameter, consisting of data sparsity, and distance from cluster head to normal nodes. To improve node clustering efficiency, a hierarchical data collection scheme is implemented, by which nodes in different layers are adaptively clustered, and the UAV can be scheduled to perform energy-efficient data collection. Simulation results show that our method can effectively collect the data and plan the path for the UAV at a low energy cost.

Index Terms—Agricultural monitoring system, artificial intelligence, industrial wireless sensor network (IWSN), intelligent signal processing, unmanned aerial vehicles (UAVs).
Thus, UAVs open a new path for gathering the data from the sensors in IWSN, upgrading the IWSN to UAV-aided IWSN, i.e., UAV-IWSN. Instead of recording data based on active data routing, UAV units can fly through each sensing field and collect data from the sensors, thus minimizing the infrastructure needed for implementing intelligent farming methods. UAV-IWSNs are aimed to be deployed over the extension of latifundia, where the dedicated sensors are far from each other, but still act as coordinated data sensing components. The UAVs perform specific data collection missions by flying over planned paths. Utilizing UAV to collect data benefits from the following advantages.

1) Compared with the vehicles on the ground, UAVs can collect data at some special places where the normal collector cannot reach to [11].
2) UAVs typically displace faster than autonomous ground vehicles, and can return to base in case of adverse weather promptly, so that the expensive equipment is not damaged.
3) UAVs are normally equipped with a high-performance computing unit and large battery capacity, and can sometimes provide the ability of lightweight data computation.
4) UAVs can be easily adapted for tasks they were not designed for, such as the detecting the incidence of wildfires and rural crime.
5) UAVs do not interfere with the livestock.

A challenge of UAV-IWSN is the deployment costs. As a consequence, not all sensor units may be equipped with all the available sensing and computing capacities. It may be more beneficial to distribute the capabilities optimally among the units, so that some units may have high-performance computing or battery-UAV communication, while others just perform the task of basic data collectors [12]. Hence, the concept of node clustering can be employed to improve the applicability and energy efficiency of UAV-IWSNs, where the cluster head (CH) is in charge of gathering the sensed data, and then transferring them to the UAV—while the other units are only performing data sensing tasks. Further, to improve the data gathering performance, the signal processing efficiency needs to be considered [8].

Succinctly, in order to perform energy-efficient data collection in UAV-IWSN, an intelligent data collection scheme integrating a node clustering algorithm, intelligent signal processing methods together with path planning policies is advantageous. With this goal in mind, in this article, the compressed sampling (CS) technique [13], [14] is employed, and we propose a hierarchical data collection scheme for an UAV-IWSN-based agricultural monitoring system. Our contribution can be summarized as follows.

1) We introduce the concept of cluster-based CS and propose a hybrid CS-based node clustering model. Based on the model, we propose a hierarchical data collection scheme integrating an exact approach and a greedy approach.
2) For the exact approach, an improved version of the minimum cost flow problem—to model the energy-optimality—is utilized. This can be formulated as an improved linear programming (LP) method.
3) In the greedy approach, we introduce a balance factor parameter, accounting for both, data sparsity and the distance from the CH to the sensing nodes.
4) The proposed scheme is tuned through an ant colony optimization (ACO) algorithm-based path planning policy.

The remainder of this article is organized as follows. Section II surveys the related work. Section III introduces the problem statement of this work. Section IV presents our proposed approach, while Section V shows some evaluation results for demonstrating and validating the scheme. Finally, Section VI concludes this article.

II. RELATED WORK

In this section, a review of the current research progress in IoT and IIoT-based agricultural monitoring systems is provided. As stated, the concept of IoI is becoming ubiquitous in a wide range of industrial environments. For agricultural applications, the motivation toward using IIoT technology is the need to embed industrial sensors into farming spaces—to monitor the irrigating system, crop, weather, soil information. For instance, in [15] the authors used IWSN technology to design crop field monitoring systems; by considering three categories of industrial sensors that survey temperature, humidity, and crop images. By analyzing these parameters, the informed decisions for maintaining the crop’s health can be made. Their proposal also contains a remote monitoring/controlling platform, more suitable for areas where the water is scarce. In [16], the authors proposed an intelligent agriculture monitoring system based on GSM, to enable automatic deployment of some important agriculture missions. By way of a GSM module, the centralized data analysis system can gather the sensed data from the sensors, and transmit the analysis result to the irrigating system to perform automatic irrigation. Their proposal is very significant to the design of cellular network-based or online agricultural monitoring systems. In [17], the authors focused on a real-time and clock-shared rainfall agricultural monitoring system for IWSN, which also served the purpose of protecting the crop from animal attacks.

In [18], novel sensors for monitoring the leaf area index (LAI) of the crops were introduced. Due to the advantages of energy-efficient, scaled-down, and low cost sensor technology, the theoretical principles behind the IEEE 802.15.4 protocol toward IWSN can be deployed in a realistic scenario, such as a large-scale cornfield. In particular, the authors showcased a project for designing a feasible, low-cost modification of commercial off-the-shelf photosynthetically active radiation sensors, which can be deployed in the LAI monitoring system. In [19], the authors proposed a clustering routing algorithm based on Dijkstra algorithm (C.R.D.A). We note that, in the clustering phase, the C.R.D.A divides the clusters by following the ELBOW method, and utilizes K-means to allocate each node to the cluster, according to the length from the normal nodes to the CHs, through a greedy approach. Their work gives a compressing clustering algorithm from the aspect of relative distance between the normal nodes and the CHs, ignoring the transmitted
data among the nodes. In [20], to measure the long-term evolutionary trend of daily average soil temperature within a specified period, the authors introduced an IWSN-based monitoring system to acquire the spatio-temporal variation of daily soil temperature. However, the maximum monitoring duration or the techniques to improve the lifetime of the monitoring network were not highlighted. In [21], the authors, respectively, utilized RFID, QR code technology, and cameras to build an IWSN-based irrigation facilities management system. Their proposal can provide extremely rewarding information to assist the facilities manager with irrigation policy determination. In [22], the authors focused on the analysis of the agronomic variables of the cassava crops, and proposed an IWSN technology-based agricultural system. In particular, each sensor in IWSN was equipped with advanced soil moisture and temperature modules, and the network performance is capable of automatic checking by evaluating the received signal strength indicator, link quality indicator, and network convergence time. In [23], low-altitude remote sensing, biosensors, chlorophyll meter, multispectral/near-infrared camera, and fluorescence spectrometer were used to build an IWSN-based monitoring system to study the relative canopy chlorophyll content in citrus orchards—by generating spatial distribution maps. It should be noted about that the authors utilized the univariate and multiple linear regressions together with the partial least squares method to propose a leaf soil and plant analyzer development value predictive model. In [24], the authors focused on the energy consumption efficiency issue and propose an adaptive energy consumption model for the IWSN-based agricultural monitoring system. Their proposal can be referred to compute the optimal number of the nodes in the network, while the connectivity and coverage of the entire network are guaranteed.

Recently, some scholars have proposed the use of UAV to assist with the agricultural information collection. For instance, UAVs can be deployed to obtain spectral or infrared images of plants to perform farmland of large scale monitoring or hazard/irrigation prediction. In [25], the authors utilized UAVs equipped with infrared and visible sensors to perform remote water stress monitoring, and compared its monitoring information with the data collected from the sensors on the ground, such that joint analysis result can be achieved. In [26], the authors proposed a agricultural monitoring system integrating IWSN and solar powered UAV to monitor CO₂ and CO₄ at the greenhouse. The UAV measures the data by hovering above the monitoring field, while the sensors of IWSN are collecting the data on the ground, leading to a 3-D monitoring environment.

To further extend the applicability of UAV-IWSN-based agricultural monitoring systems, we introduce the data-driven node clustering policy especially from the aspect of data CS, enabling an UAV-assisted data collection framework.

**III. Problem Statement**

### A. Network Model

To enable intelligent, automatic, and informative industrial agriculture, we adopt an IoT framework perspective to introduce a scheme for an UAV-IWSN-based intelligent farm. In Fig. 1, the proposed UAV-IWSN is shown, consisting of an UAV and a series of wireless monitoring sensors divided by different clusters. The sensors are deployed for different agricultural monitoring purposes, e.g., soil moisture/acidity, temperature, air components, energy status, etc. Instead of using active data routing to collect and monitor the data at the base station, a preconfigured rechargeable UAV is scheduled to fly through each sensing field. Then, the UAV uploads the data to the data center, where the data are automatically computed/analyzed, and the agricultural management policy is adaptively adjusted, e.g., irrigation levels and adequate feed nutrient content.

As discussed, in the proposed network architecture, we divide the sensors into different clusters, each of which is managed by a CH. The q normal nodes of the UAV-IWSN are denoted as v₁, v₂, ..., v₉, while the p CHs as ch₁, ch₂, ..., chₚ—which are related to clusters the C₁, C₂, ..., Cₚ, respectively. We utilize the notation

\[ S(v_i) = C_x \]  \hspace{1cm} (1)

and

\[ W(C_x) = \{ v_i | S(v_i) = C_x \} \]  \hspace{1cm} (2)

for the number of CH accounts for a very small percentage.

1) The CH is equipped with high-performance battery, data storing, communication components, and is more expensive than the normal node. As a result, in the proposed network architecture, the number of CH accounts for a very small percentage.

2) Each normal node is assigned to a dedicated cluster for performing data collection, under the management of the CH. To improve the energy consumption efficiency, all the normal nodes forward their collected data to the CHs, instead of communicating directly with the UAV.

3) The CHs store the gathered data and wait for the data uploading scheduling. The UAV performs nonreal-time
data collection from the CHs, based on a preconfigured data collection policy.

4) The features (e.g., the location, the CH of each node) of each cluster are kept constant once the clusters are determined and deployed.

In a cluster $C_x$, the communication range among the sensors is imposed by a threshold $r$ (determined when both of the energy consumption, packet loss rate, etc., are concurrently taken into account). And, the data can be forwarded to the $ch_x$ by utilizing multihop routing algorithm based on the spanning tree $G_x(V_x, E_x)$ derived from $C_x$. In $G_x(V_x, E_x)$, $V_x = W(C_x) \cup ch_x$, and $e_{ij} = \langle v_i, v_j \rangle \in E_x$ is a corresponding edge if the distance between $v_i$ and $v_j$ does not exceed $r$.

### B. Energy Consumption Model

We refer to the energy consumption model in [27] and respectively use the following to compute the energy consumption, when $n$ bits of data are sent and received:

$$C_s(n, r) = \begin{cases} n(\theta_{fs}r^2 + C_a) & r < R \\ n(\theta_{mp}r^4 + C_a) & r \geq R \end{cases}$$

(3)

$$C_r(n) = nC_a.$$  

(4)

In (3), both $\theta_{fs}$ and $\theta_{mp}$ represent the sending power coefficients; $\theta_{fs}r^2$ and $\theta_{mp}r^4$, respectively, denote the consumed energy by the transmit amplifier for sending one bit of data within communication range $r$, when the transmit amplifier works in different modes. We notice that this mainly depends on both the communication range and the accepted bit error bit; $R$ denotes the distance threshold in the free space model. In (3) and (4), $C_a$ denotes the consumed energy to activate the sending or receiving circuit. Both of the aforementioned variables are determined based on the real features of the physical electronic components.

### C. Problem Definition

Given an UAV-IWSN, represented as graph $G(V, E)$, that is deployed in an open area of an intelligent farm, where $V$ is the set of nodes (including both normal nodes and CHs) and $E$ is the potential edge set. Based on the aforementioned network model in Section III-A, the proposed optimal data collection (ODC) problem aims at defining a suitable clustering method, together with a signal acquisition approach, to minimize the total energy consumption. Then, the ODC problem also requires to find a path planning algorithm to schedule the data gathering path for the UAV, such that the energy consumption for UAV is minimized.

The studied ODC problem is very common in real industrial agricultural monitoring systems, due to the wide deployment of the monitoring sensors, the un rechargeable features of the sensors, etc.

### IV. DATA COLLECTION SCHEME

#### A. Proposed Hybrid CS-Based Cluster Model

In this article, to improve the data collection efficiency and decrease the energy consumption, we employ the CS technology and propose to utilize the novel hybrid CS approach to address the signal acquisition problem. By hybrid CS technology, in a cluster, the normal sensing nodes can be divided into uncompressed and compressed nodes. In particular, as the intermediate nodes, the compressing nodes cannot only transmit the collected data following the routing algorithm, but also compress the data traversing them. To clearly explain our idea, we use the instance in Fig. 2 to clarify our cluster-based data compression model.

In this article, we focus on the optimal energy consumption-based clustering problem based on the model in Section III-B. Therefore, we ignore the procedure of data compressing.

#### B. Exact Approach

For a given cluster $C_x$, $\gamma_x$, $D_x$ denote the data compression ratio of $C_x$ and the scale of the gathered data, respectively. Thus, the scale of the compressed data can be expressed by

$$\sigma_x = \frac{D_x}{\gamma_x}.$$  

(5)

According to the depiction for the ODC problem, we can infer that different node clustering schemes will result in different spanning trees with different compression ratios. In the following, we will present the exact expression for addressing the ODC problem.

Let $e_{ij}$ be a potential link between the nodes $v_i$ and $v_j$ in the spanning tree $G_x(V_x, E_x)$, and $n_{ij}$ the data to be transmitted over $e_{ij}$. Following (3) and (4), the energy costs for transferring
\[ n_{ij} \] units of data over link \( e_{ij} \) can be computed as
\[
C(e_{ij}) = C_s(n_{ij}, d_{ij}) + C_r(n_{ij})
\]
while the energy consumption for \( C_x \) is expressed as
\[
C(C_x) = \sum_{e_{ij} \in E_x} C(e_{ij}).
\]
Thus, the entire consumed energy for the entire network is
\[
C_{all} = \sum_{x=1}^{p} C(C_x).
\]
By following the hybrid CS theory, we acquire the following LP formulation to express the optimization model.
\[
\min_{C_x} C_{all} = \sum_{x=1}^{p} C(C_x)
\]
\[
\text{s.t. } \ 
\sum_{j : e_{ij} \in E_x} n_{ij} \geq \sum_{k : e_{ik} \in E_x} (\delta |C_x| - \sigma_x) + \delta, \forall v_i \in C_x
\]
\[
\sum_{j : e_{ij} \in E_x} n_{ij} \geq \sum_{k : e_{ik} \in E_x} (\sigma_x - \delta) \xi_i + \delta, \forall v_i \in C_x
\]
\[
\sum_{j : e_{ij} \in E_x} m_{ij} \geq \sum_{k : e_{ik} \in E_x} m_{ki} + \frac{1}{|C_x|}, \forall v_i \in C_x
\]
\[
m_{ij} \geq m_{ij}, \forall e_{ij} \in E_x
\]
\[
\sum_{e_{ij} \in E_x} n_{ij} \geq 1
\]
\[
m_{ij} \geq \sigma_x - n_{ij}, \forall e_{ij} \in E_x
\]
\[
S(v_i) = C_x, i = 1, 2, \ldots, q; x = 1, 2, \ldots, p
\]
\[
W(C_x) = \{v_i | S(v_i) = C_x\}
\]

In (9), \( m_{ij} \in \{0, 1\} \) indicates whether \( e_{ij} \) in \( G_x(V_x, E_x) \) represents the last link connecting a normal node and \( ch_x \); \( \xi_i \in \{0, 1\} \) is an indicator variable for an intermediate node \( v_i \); \( |C_i| \) denotes the number of the nodes in cluster \( C_i \); \( m_{ij} \) or \( m_{ki} \) denotes the flow over visual link (indicating the connectivity of two adjacent nodes). Equation 9 is actually an extension of the minimum-cost flow problem. The optimization in (9) is aimed at seeking node clustering together by way of spanning tree construction, i.e., \( C, x = 1, 2, \ldots, p, \) to minimize the entire energy consumption. In particular, the constraints present the rules for building \( G_x \) for \( C_x, x = 1, 2, \ldots, p \). More in detail, the first two constraints guarantee the flow conservation at the intermediate node, where the data CS occurs. Constraints three to seven ensure the flow over each link in the network is nonnegative. The last two constraints prevent the nodes \( (v_i, i = 1, 2, \ldots, q) \) from belonging to more than one of the clusters \( C_x, x = 1, 2, \ldots, p \).

Furthermore, from (9), we can see that the proposed exact approach is an integer LP-based solution, that cannot be run in polynomial time. In the following section, we will present a greedy approach—an approximate optimal solution, to efficiently address the ODC problem.

C. Greedy Approach

From (5), the smaller the \( \gamma_x \), the larger the required \( D_x \), i.e., more data collection is required. As a result, more data require to be forwarded to the CH, resulting in more energy consumption. Furthermore, following (3), larger data transfer distances result in the increase of energy consumption. Therefore, the energy consumption of cluster \( C_x \) depends on two factors: the compression factor \( \gamma_x \) and the distance \( r_{ix} \) between the normal node \( v_i \) and \( ch_x \).

To improve the energy consumption efficiency, \( r_{ix} \) should be minimized while improving the compression ratio \( \gamma_x \) of cluster \( C_x \), which involves a multiobjective tradeoff. To allocate \( v_i \) into an approximate optimal cluster, we define the balance factor \( B(i, x) \), which quantify the effect of allocating \( v_i \) into a existing cluster \( C_x \)
\[
B(i, x) = B_s(i, x) \rho + B_c(i, x)(1 - \rho)
\]
where \( B_s(i, x) \) is to compute the difference of data sparsity when \( v_i \) is allocated to cluster \( C_x \), \( B_c(i, x) \) is to compute the normalized distance between \( v_i \) and \( C_x \), and \( \rho \in [0, 1] \) aims at specifying the weight of \( B_s(i, x)/B_c(i, x) \) in \( B(i, x) \).

In particular, \( B_s(i, x) \) and \( B_c(i, x) \) can be, respectively, computed by
\[
B_s(i, x) = \frac{DD(i, x)}{\max\{|DD(:, :)|\}}
\]
\[
B_c(i, x) = \max\{r_{\alpha \beta}, \alpha = 1, 2, \ldots, q, \beta = 1, 2, \ldots, p\}
\]

In (11), \( DD(i, x) \) represents the difference of data sparsity based on discrete cosine transform (DCT) algorithm and data CS, and \( \{|DD(:, :)|\} \) denotes the set of the difference value when \( v_i \) is allocated to different clusters. In (12), \( \{r_{\alpha \beta}, \alpha = 1, 2, \ldots, q, \beta = 1, 2, \ldots, p\} \) represents the set of distance value between each \( v_i \) and each \( ch_x \).

In (10), \( \rho \) specifies the slope of \( v_i \), when \( v_i \) is allocated to a cluster \( C_x \). The higher the value of \( \rho \), the more inclined \( v_i \) is to be allocated to the cluster with a higher compression ratio, else the cluster that is closer to \( v_i \) is more inclined.

After \( \rho \) is determined, the \( C_x \) that minimizes \( B(i, x) \) is selected as the candidate cluster. Thus, this leads to a greedy approach which can cluster one node at a time, until the clustering problem of the ODC task addressed. However, as an approximate optimal scheme, the optimal approach based on computing \( B(i, x) \) can sometimes acquire multiple locally optimal solutions.

In the following section, we propose a hierarchical data collection scheme based on a proposed hybrid clustering scheme, integrating both the proposed exact and greedy approaches.

D. Hybrid Clustering-Based Data Collection Scheme

In the proposed UAV-IWSN, all the computing operations (e.g., data sparsity computation based on DCT algorithm, constructing spanning tree, etc.) are performed onboard in the UAV, since it is equipped with high-performance computing and data storing capabilities. We propose Algorithm 1, a hybrid optimal approach, to perform node clustering. In line 1 of Algorithm 1,
Algorithm 1: Node Clustering($V_u$).

1: Initialize $V_u$.
2: for $v_i \in V_u$ do
3:  for $C_x, x = 1, 2, \ldots, p$ do
4:  Following Eq. (10), Eq. (11), and Eq. (12), computing $r_{ix}$, $DD(i, x)$, $\max\{DD(:, i)\}$, and $\max\{r_{\alpha\beta}, \alpha = 1, 2, \ldots, q, \beta = 1, 2, \ldots, p\}$, respectively.
5:  Compute $B_s(i, x)$, $B_c(i, x)$, and $B(i, x)$.
6:  end for
7:  if $|c_{h_{\min}}| \geq 2$ then
8:    Allocating $v_i$ to the $C_m$ with $c_{h_{\min}} \in c_{h_{\min}}$, leading (9) to be minimized.
9:  end if
10: end for
11: Return the computed node clustering policy.

$V_u$ represents the set of unclustered nodes. Algorithm 1 executes node clustering individually for $v_i \in V_u$, as shown in lines 2 to 11; first, lines 3–7 compute a set $c_{h_{\min}}$ of optimal solution greedily. In particular, the LEACH algorithm [28] is invoked to execute the node clustering for data collection in the first round, such that the data compression ratio of each cluster can be computed. Then, in line 9, the proposed exact approach is utilized to select a solution that can minimize (9).

From Algorithm 1, we can see that the DCT algorithm requires to be frequently invoked, which takes most of the running time. By (10), $v_i \in V_u$ is preferentially allocated to $C_x$, when $r_{ix}$ is very small, or smaller than a given threshold.

With this property, in the following Algorithm 2, we introduce a hierarchical data collection scheme to improve the data collection efficiency. In Algorithm 2, after the initialization phase in lines 1–2, line 4 constructs the sequence $S_{q_i}$ for $v_i$. Note that the sequence is in the ascending order of $r_{ix}, x = 1, 2, \ldots, p$. For instance, Fig. 3(a) shows the construction for the sequence $S_{q_i}$ of $c_{h_x}$ for $v_i$. Then, following line 5 of Algorithm 2, allocate $v_i$ to the cluster whose CH is marked in $S_{q_i}[0]$, i.e., $C_3$ in Fig. 3(a). After that, we present a hierarchical node classification scheme shown in lines 8–9. Line 8 groups the nodes in $C_x$ into $h$ layers of equal range from 1 to $r_{max}$ [see (13)]. For instance, in Fig. 3(b), the nodes in $C_x$ are divided into $h$ parts: $C_x^1, C_x^2, \ldots, C_x^h$. By specifying $\mu$ based on the practical scenario, allocating the nodes in $C_x^1, C_x^{h-1}, \ldots, C_x^h (1 < \mu < h)$ to the node set $V_u$. This leads to a node reclustering phase for the nodes in $V_u$ by invoking Algorithm 1, as shown in line 11 of Algorithm 2. Thus, the running efficiency of the node clustering phase will be improved. Then, line 12 constructs $G_x(V_x, E_x), x = 1, 2, \ldots, p$ by evoking MECDA_GREEDY algorithm [29], and the data in each cluster follow the specified hybrid CS policy and is forwarded along the spanning tree to the CH. In this article, we assume the consumed power of the UAV is proportional to the path length, take the shortest possible route (that the UAV starts from the source, traverses each CH, and then goes back to the source) into account, and treat the path planning optimization problem as a travelling salesperson problem. Hence, in line 13, the ACO algorithm [30] can be selected as a candidate approach, to schedule the data collection path for the UAV when the energy consumption efficiency is taken into account.

$$r_{max} = \max\{r_{ix}, i = 1, 2, \ldots, q\}. \quad (13)$$

V. EVALUATIONS

In this section, we present simulation results to evaluate the performance of the proposed approach for tackling the ODC problem. We first test the performance of the proposed hierarchical clustering algorithm, and then we present some comparison results for demonstrating our proposal. Further, we survey some normal path planning schemes to demonstrate the
data collection efficiency for the UAV, when the proposed data collection scheme is performed.\(^1\)

The simulations were implemented in Python 3.7 over a 2-D array representing a 2000 m\(^2\) area, and where 100 nodes with \(p\) CHs were randomly distributed. The communication among the nodes follows the Zigbee protocol [31], and the scale of each frame is limited to be 108 B. The forwarded data on each node is in float (4 B for each unit). All the clusters are with the same compression ratio. The simulation parameters are summarized in Table I.

![Table I: Parameter Setting During Model Evaluation](image)

| parameter       | description                                      | value                        |
|-----------------|--------------------------------------------------|------------------------------|
| \(|V|\)          | the number of nodes                              | 100                          |
| \(\theta_{fs}\) | sending power coefficient                        | 11 pJ/bit/m\(^2\)            |
| \(\theta_{imp}\)| sending power coefficient                        | 0.00145 pJ/bit/m\(^2\)       |
| \(C_a\)        | consumed energy to activate the sending or receiving circuit | 65 nJ/bit                   |
| \(r\)          | communication range of each node                 | 200 m                       |
| \(p\)          | number of cluster                                | \([5, 6, \ldots, 24]\)      |
| \(h\)          | number of layer in total                         | \([5, 6, \ldots, 24]\)      |
| \(\mathcal{R}\) | distance threshold in the free space model       | 240 m                       |
| \(\mu\)        | number of layer leading to a node re-clustering phase | \([5, 6, 7, 8, 9]\)        |
| \(n_{ij}\)     | scale of the forwarded data on each node         | 40                          |
| \(\gamma_{x}\) | compression ratio of each cluster                | 50                          |

\(^1\)The simulations were conducted on an Intel(R) Core i7-8565 U 1.8 GHz machine with 8 Gb RAM.

![Fig. 4. Test for the studied scheme. (a) Energy consumption with \(h/\mu\). (b) Running time with \(h/\mu\). (c) Ratio of multiple optimal solutions with \(p\).](image)

![Graphs showing energy consumption, running time, and ratio of multiple optimal solutions.](image)

are repetitions executed, when the nodes are divided into 5–24 clusters. The ratio of acquiring multiple locally optimal solutions is shown in Fig. 4(c); we can see that the ratio increases with the number of clusters, i.e., a larger number of clusters the greater the number of potential options.

Furthermore, to evaluate our approach, we also make comparisons with the other four schemes: DSNO, DDS, EXNO, HNO, and the scheme proposed in [19]. In particular, DSNO is the scheme that does not take the data compression into account; DDS is a scheme that only considers the distance between the normal node and CHs; EXNO is an approach that ignores the exact approach-based decision procedure; HNO is a method that is irrespective of the hierarchical node clustering procedure. In particular, we refer to the clustering scheme in [19] as the k-CL which ignores as well the data compression. We first test the energy consumption for the various algorithms. The result in Fig. 5(a) demonstrates that the proposed scheme benefits from the hierarchical hybrid clustering scheme and displays improved performance as compared to the other methods. In addition, we can see that the data compression skill in the proposed hierarchical hybrid clustering scheme has a maximum impact on the energy consumption result during our evaluation. Furthermore, the execution time results in Fig. 5(b) demonstrate that the proposed hierarchical node clustering framework can
improve the running efficiency of the algorithm, and the exact approach-based decision procedure takes most of the running time in the proposed scheme.

Finally, we utilize our proposal to plan the path for the data collecting UAV after the monitoring nodes are clustered. To understand why the ACO algorithm is selected to plan the path for the UAV in our proposed scheme, we compare ACO algorithm-based approach with the other two normal schemes: the simulated annealing (SA) algorithm [32] and the self-organizing-map (SOM) algorithm-based path planning schemes [33], especially in the aspect of path length and running efficiency. We select the coordinates in the first footnote as the CH coordinates that are distributed in a 8000 m² area.

In Fig. 6, we select 12–48 CHs and, respectively, use the aforementioned schemes to plan the data collecting path and compute the path length together with the running time. In particular, all the results in Fig. 6 are the mid-values based on 1000 runs. As the test results for the path planning length shown in Fig. 6(a), the ACO algorithm-based path planning scheme performs the best and is closest to the optimal solution. Meanwhile, in Fig. 6(b), we also note that the running time of the ACO algorithm-based path planning scheme increases with the number of the CH and performs the best when the number of the CH is less than 18. From Fig. 6(b), we can as well see that the performance of ACO is between SA and SOM. To summarize, we select ACO algorithm as the candidate approach to plan the data collecting path for the UAV in our scheme.

Besides, in Fig. 7, we also show a path planning case with 48 CHs based on the ACO algorithm-based approach. Assuming the data collection is started from CH 24, the path along the red arrow, i.e., $<24, 10, 42, 5, 48, 39, 32, 21, 13, 25, 14, 23, 11, 12, 33, 46, 15, 40, 9, 1, 8, 38, 31, 44, 18, 7, 28, 36, 6, 37, 19, 27, 43, 17, 30, 20, 47, 3, 22, 16, 41, 34, 29, 2, 26, 4, 35, 45>$ can be selected as an approximate-optimal path. The result in Fig. 7 demonstrates as well that our scheme can accurately guarantee an efficient data collection with no duplicate path, leading the ODC problem to be accurately addressed.
VI. CONCLUSION

In this article, we employed the paradigm of intelligent agriculture to improve farming automation digitalization, and proposed a hierarchical data collection scheme to perform energy-ODC in UAV-IWSN-based agricultural monitoring systems. To improve the energy efficiency of data gathering from the UAV-IWSN nodes, the concept of data CS and node clustering was introduced. For that, a hybrid compressed technique for clustering was implemented. Based on the proposed model, we introduced a hierarchical data collection scheme integrating an exact approach and a greedy approach. By dividing the nodes into different layers, the exact approach and greedy approach can be intelligently matched. Particularly, for the exact approach, an improved version of the minimum cost flow problem—that can be expressed by LP formulation to model the energy-optimal problem—is designed. The greedy approach was based on a balance factor parameter, consisting of both, data sparsity and the distance from the CH to the normal nodes. The proposed scheme was tuned with an ant-colony-optimization path planning policy. Simulation results showed that this method can efficiently gather the data from several normal schemes, especially in energy consumption, and plan the path for the UAV at a low energy cost.

For future work, we plan to expand our work from the aspect of network architecture optimization. For instance, we intend to utilize software-defined networking (SDN) technology to improve the scalability of the monitoring network, such that the network status can be intensively monitored and surveyed, and the network operation can be uniformly deployed.

REFERENCES

[1] A. Antonacci, F. Ardini, D. Moscone, G. Palleschi, and V. Scognamiglia, “Nanostructured (bio)sensors for smart agriculture,” TrAC Trends Analytical Chem., vol. 98, no. 1, pp. 95–103, Jan. 2018.

[2] S. Heble, A. Kumar, K. V. V. D. Prasad, S. Samirana, P. Rajalakshmi, and U. B. Desai, “A low power IoT network for smart agriculture,” in Proc. IEEE World Forum Internet Things, Coimbatore, India, 2018, pp. 609–614.

[3] O. Westermann, W. Freh, P. Thornton, J. Krner, L. Cranmer, and B. Campbell, “Scaling up agricultural interventions: Case studies of climate-smart agriculture,” Agricultural Syst., vol. 165, no. 9, pp. 283–293, Sep. 2018.

[4] M. Martinez-Garcia, Y. Zhang, K. Suzuki, and Y.-D. Zhang, “Deep recurrent entropy adaptive model for system reliability monitoring,” IEEE Trans. Ind. Informat., to be published, doi: 10.1109/TII.2020.3007152.

[5] R. Alcarria, B. Bordel, M. Á. Manso, T. Iturrioz, and M. Pérez, “Analyzing UAV-based remote sensing and WSN support for data fusion,” in Proc. Int. Conf. Inf. Technol. Syst., 2018, pp. 756–766.

[6] M. Aazam, K. A. Harass, and S. Zeadally, “Fog computing for 5G tactile industrial Internet of Things: QoE-aware resource allocation model,” IEEE Trans. Ind. Informat., vol. 15, no. 5, pp. 3085–3092, May 2019.

[7] M. H. ur Rehman, I. Yaqoob, K. Salah, M. Imran, P. P. Jayaraman, and C. Perera, “The role of Big Data analytics in industrial Internet of Things,” Future Gener. Comput. Syst., vol. 99, no. 10, pp. 247–259, Oct. 2019.

[8] G. Han, X. Miao, H. Wang, L. Liu, J. Jiang, and Y. Peng, “A dynamic multipath scheme for protecting source location privacy using multiple sinks in WSNs intended for IloT,” IEEE Trans. Ind. Informat., vol. 16, no. 8, pp. 5527–5538, Aug. 2020, doi: 10.1109/TII.2019.2953937.

[9] L. Liu, G. Han, Y. He, and J. Jiang, “Fault-tolerant region detection on trajectory pattern extraction for industrial wireless sensor networks,” IEEE Trans. Ind. Informat., vol. 16, no. 3, pp. 2072–2080, Mar. 2020, doi: 10.1109/TII.2019.2933238.

[10] M. Martinez-García, Y. Zhang, and T. Gordon, “Modeling lane keeping by a hybrid open-closed-loop pulse control scheme,” IEEE Trans. Ind. Informat., vol. 12, no. 6, pp. 2256–2265, Oct. 2016.

[11] D. Popescu, F. Stoican, G. Stamatescu, O. Chenaru, and L. Ichim, “A survey of collaborative UAV–WSN systems for efficient monitoring,” Sensors, vol. 19, no. 21, Oct. 2019, Art. no. 4690.

[12] S. Wang, Q. Zhou, and C.-T. Cheng, “A UAV-assisted topology-aware data aggregation protocol in WSN,” Phys. Commun., vol. 34, no. 7, pp. 48–57, Jun. 2019.

[13] Q. Wang, D. Lin, P. Yang, and Z. Zhang, “An energy-efficient compressive sensing-based clustering routing protocol for WSNs,” IEEE Sensors J., vol. 19, no. 10, pp. 3954–3962, May 2019.

[14] V. K. Singh, M. Kumar, and S. Verma, “Node scheduling and compressed sampling for event reporting in WSNs,” IEEE Trans. Netw. Sci. Eng., vol. 6, no. 3, pp. 418–431, Jul. 2019.

[15] P. Rajalakshmi and S. Devi Mahalakshmi, “IoT based crop-field monitoring and irrigation automation,” in Proc. IEEE Int. Conf. Intell. Syst. Control, Coimbatore, India, 2016, pp. 1–6.

[16] C. Dwarkani M. G. Ram R. S. Jagannathan, and R. Priyatharshini, “Smart farming system using sensors for agricultural task automation,” in Proc. IEEE Technol. Innov. ICT Agriculture Rural Devol., Chennai, India, 2015, pp. 49–53.

[17] G. Sun, T. Hu, G. Yang, and J. Jia, “Real-time and clock-shared rainfall monitoring with a wireless sensor network,” Comput. Electron. Agriculture, vol. 119, no. 11, pp. 1–11, Nov. 2015.

[18] J. Bauer, B. Siegmann, T. Jarmer, and N. Aschenbruck, “On the potential of wireless sensor networks for the in-situ assessment of crop leaf area index,” Comput. Electron. Agriculture, vol. 128, no. 10, pp. 149–159, Oct. 2016.

[19] M. Abderrahim, H. Hakim, H. Boujemaa, and F. Touati, “A clustering routing based on Dijkstra algorithm for WSNs,” in Proc. IEEE Int. Conf. Sci. Techn. Autom. Control Comput. Eng., 2019, pp. 605–610.

[20] L. Hui, M. Zhijun, W. Hua, and X. Min, “Spatio-temporal variation analysis of soil temperature based on wireless sensor network,” Int. J. Agricultural Biol. Eng., vol. 9, no. 6, pp. 131–138, Jun. 2016.

[21] W.-H. Nam, T. Kim, E.-M. Hong, J.-Y. Choi, and J.-T. Kim, “A wireless sensor network (WSN) application for irrigation facilities management based on information and communication technologies (ICTs),” Comput. Electron. Agriculture, vol. 143, no. 12, pp. 185–192, Dec. 2017.

[22] J. G. Caicedo-Ortiz et al., “Monitoring system for agronomic variables based in WSN technology on cassava crops,” Comput. Electron. Agriculture, vol. 145, no. 2, pp. 275–281, Feb. 2018.

[23] K. Wang, W. Li, L. Deng, Q. Lyu, Y. Zheng, and S. Yi, “Rapid detection of chlorophyll content and distribution in citrus orchards based on low-altitude remote sensing and bio-sensors,” Int. J. Agricultural Biol. Eng., vol. 11, no. 2, pp. 164–169, Feb. 2018.

[24] D. L. Ndzi et al., “Wireless sensor network coverage measurement and planning in mixed crop farming,” Comput. Electron. Agriculture, vol. 105, no. 7, pp. 83–94, Jul. 2014.

[25] J. Gago et al., “UAVs challenge to assess water stress for sustainable agriculture,” Agricultural Water Manage., vol. 153, no. 5, pp. 9–19, May 2015.

[26] A. Malaver, N. Motta, P. Corke, and F. Gonzalez, “Development and integration of a solar powered unmanned aerial vehicle and a wireless sensor network to monitor greenhouse gases,” Sensors, vol. 15, no. 2, pp. 4072–4096, Feb. 2015.

[27] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in Proc. IEEE Annu. Hawaii Int. Conf. Syst. Sci., Maui, HI, USA, USA, 2000, pp. 1–10.
Chuan Lin (Member, IEEE) received the B.S. degree in computer science and technology from Liaoning University, Shenyang, China, in 2011, the M.S. degree in computer science and technology and the Ph.D. degree in computer architecture from Northeastern University, Shenyang, China, in 2013 and 2018, respectively.

He was a Visiting Ph.D. Student with Network Architecture and Services Group, Delft University of Technology, Delft, the Netherlands, during 2014–2015. He is currently an Associate Professor with Software College, Northeastern University. His current research interests include UWSNs, industrial IoT, software-defined networking.

Guangjie Han (Senior Member, IEEE) received the Ph.D. degree in computer science from Northeastern University, Shenyang, China, in 2004.

From 2004 to 2005, he was a Product Manager with ZTE Company, Shenzhen, China. From 2005 to 2006, he was a Key Account Manager with the Huawei Company, Shenzhen, China. In February 2008, he was a Postdoctoral Researcher with the Department of Computer Science, Chonnam National University, Gwangju, South Korea. From January 2017 to February 2017, he was a Visiting Professor with the City University of Hong Kong, Hong Kong. He is currently a Professor with the Department of Information and Communication System, Hohai University, Changzhou, China, and a part-time researcher with Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, 350118, China. He has authored over 400 papers published in related international conference proceedings and journals is the holder of 125 patents.

Jiaxin Du received the B.S. degree from the Taiyuan University of Science and Technology, Taiyuan, China, in 2016. She is currently working toward the Ph.D. degree at the Dalian University of Technology, Dalian, China, both in software engineering.

Her current research interests include privacy protection and security for underwater wireless sensor networks.

Tiantian Xu received the M.S. degree in software engineering from the Tianjin Normal University of Science and Technology, Tianjin, China, in 2019. She is currently working toward the Ph.D. degree in electronic and information at the Dalian University of Technology, Dalian, China.

Her current research interest includes fault detection of Industrial Internet of Things.

Xingyue Qi received the B.S. degree from GuiZhou University, Guiyang, China, in 2019. He is currently working toward the M.S. degree at the Dalian University of Technology, Dalian, China, both in software engineering.

Her current research interests include intelligent fog computing, SDN, and underwater multi-AUV cooperation for data collection.

Miguel Martinez-Garcia received the B.Sc. degree in mathematics and the M.Sc. degree in advanced mathematics and mathematical engineering (MAMME) from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 2013 and 2014, respectively, and the Ph.D. degree in engineering from the University of Lincoln, Lincoln, U.K., in 2018.

Since 2017, he also worked as a Researcher with the University of Lincoln and the Advanced Virtual Reality Research Centre (AVRRC), Loughborough, U.K. In 2020, he became a Lecturer in human-machine systems with Loughborough University, Loughborough, U.K. His current research interests include human–machine integration, machine learning, artificial intelligence, intelligent signal processing, and complex systems, with particular focus in the analysis of nonlinear signals representing phenomena of interest between humans and machines.