A Data Collection Framework for Tracking Collective Behavior Patterns

Erick Stattner, Martine Collard, Philippe Hunel, Nicolas Vidot
LAMIA, University of Antilles and Guyane, France
e.stattner@gmail.com, {mcollard, phunel, nvidot}@univ-ag.fr

***Abstract*—Many initiatives have been taken to observe the adaptation of endangered animal species to various environmental changes they face and their chances of surviving. In this paper, we propose an original solution that consists on a new framework based on wireless sensor technology that enables to track individual movements and to collect spatio-temporal data to be analyzed without any harmful contact with animals. We discuss network configuration parameters what may optimize the elicitation of two specific kinds of collective motion behavior from such recorded data.**

I. Introduction

Among the issues threatening ecosystems is the disappearance of animal species, a sign of unbalanced environment. The observation of animal populations in their original area has motivated numerous scientific experiences in which in most cases individual behaviors are manually recorded by scientists. This task is difficult especially in the case of birds since human intrusion [8], may itself cause change in their way of life and be harmful.

Mobile devices have provided new perspectives thanks to their miniaturization and technological advances both in terms of computational power and memory capacity. Since they provide regular information on their positioning in time and space, geographic locators are used to collect spatial-temporal data enabling to track animal motions remotely. This technique is conceivable nor for observing frail animals such as little birds that may greatly participate in the ecological balance and need attention, neither for tracking flock movement. While they bring a very interesting alternative to physical intrusion in animal communities areas, they nevertheless need at least a minimal human interaction that is sometimes not advisable.

In our work, we have focused on this particular issue to propose an alternative that combines wireless sensor networks to collect data and data mining techniques to process them. The objective is to record data that inform about individual positions in time and space in order to study group motions and thus improve current knowledge on a given specie. We have been particularly interested in the case of a bird, the "Moqueur Gorge Blanche", an endangered species endemic to the island of Martinique. The work we conducted could be applied on similar cases. With the objective to avoid any trouble on the bird community, we have proposed a new architecture based on fixed sensors fitted with microphones where sensors are located randomly as if they were thrown by plane. Sensors are able to detect the presence of individuals on the basis of their songs only. Spatio-temporal data deduced from song records are broadcasted through a two layer sensor network and they are analyzed for eliciting group movement patterns.

This paper focuses mainly on a first issue that is to define and configure the sensor network to be set up on the ground. In this preliminary step, our objective is to simulate various configurations in order to define optimal parameters for the real architecture. To the best of our knowledge, this kind of framework based on audio record and wireless sensor fitted with microphones has not yet been exploited to elicit flock motion patterns.

Movements are very essential for understanding the way endangered communities are organized and how they live in an area. Thus, studying movements provides important information not only on social structure and habits, but on the way these movements evolve and which elements influence them [4], [7].

The paper is organized as follows. Section II introduces the motion pattern detection problem. In Section III, we describe the data collection framework we designed and we give a more formal definition of movement patterns. Section IV and Section V are each devoted to the specific flock and periodic collective motions. Section VI details experimental results obtained by simulating the sensor networks with various parameters. Finally, in Section VII, we conclude and present our future research directions.

II. Motion Pattern Detection

Recent development of technologies for real-time tracking of moving entities (GPS, mobile phones, RFID chips, etc.) has allowed the collection and exploitation of a new type of data so called spatio-temporal data that relate to time and space. They are often used to represent and study the trajectories of entities moving in spaces with two or three dimensions. The exploitation of this kind of data to track spatio-temporal movements of groups or communities has motivated early works. However, most methods on moving entities (humans, animals or objects) assume either the observation by a human observer, or mobile devices placed on the entities.

A first study initiated in movement patterns focused on flock movement [20], [16]. A flock pattern in a time interval T can be defined informally by the movement of m entities, such as for each point taken in the time interval T, we can find the same set of m entities at a different location. For example, this problem was studied in [3] by Benkert et al.

and assumed human interaction and manual data collection. Subsequently, it was established that collective movements of entities are limited to a finite number of patterns. Let us take the example of a company's employees. Everyday, they wake up at the same time and follow and take more or less the same route to their work [21]. This repetitive character of the movement can also be found for vehicles, animals and humans. From this observation, Laube et al. [19] defined a set of spatio-temporal models based on the characteristic of movements such as flock pattern and periodic pattern. With a different point of view Gudmundsson and al. [13] distinguished new movement patterns : flock patterns [3], leadership patterns, periodic patterns [21] and meeting patterns or frequent patterns.

Various methods are proposed for detecting these patterns [18], [14] and most of them employ data mining techniques [27], including clustering [20], [16] and association rules [25], [5] that have greatly contributed to their success. Verhein [25] showed how it is possible to describe the entities movements by association rules. Similarly, Gudmundsson et al. [13] provide a very complete overview of data mining contributions in the development of algorithms for research patterns in the movement.

The low cost of micro-devices that provide spatial-temporal information about moving entities are now involved in the ethology domain. Various studies using GPS collars to collect the animals motions were conducted. For example, GPS collars were experimented by Rumble and al. [23] on elks to obtain their positions every twenty minutes and by Dumon and al. [9] on sheep for studying their collective motions. Associated with environmental data such as temperature, light or noise, spatial-temporal data can provide valuable information about the movement of endangered species and facilitate their reintegration in similar environments.
As proved by a lot of experiments, GPS systems gather several advantages. But in our context, it was not conceivable to consider such a solution for several reasons listed below. A GPS system:

- only works effectively in open areas without any object which may obstruct the field of view of the receiver. Thus, operation under dense foliage is not possible.
- is particularly expensive. For large populations of birds, it would be difficult to fit each bird with a GPS device.
- consumes energy enormously.
- may have a sizeable weight, and often cannot be placed on a small bird because of risks it cannot fly anymore, or it is tired very rapidly.
- may be harmfulness since it needs to operate not only in the animal area but on selected specimens too.

On the other hand, sensors are fixed devices in the space. Unlike GPS collars, they are not grafted onto a single individual, but simply set up on the target area. By fitting each sensor with a microphone, it is possible to detect the presence of a bird when it sings. Obviously, although the sensors do not have the same problems as the GPS devices, they also have some disadvantages that have to be considered. First of all, as explained just above, sensors detect the presence of a bird only if and when it sings. And we know that birds do not sing all the time. In addition, signals recorded by sensors do not allow identifying one individual. It only ensures to recognize the specie. Since we are only interested in collective patterns, this is not a real drawback as we show in further sections.

Other technologies like mobile phones supply different media to collect spatio-temporal data too. For instance, triangulation schemes are proposed [2] to locate, track and count mobile phones. For obvious reasons, this cannot be a solution for this experiment.

III. The Proposed Framework

In the work presented here, the experimental context set strong constraints that contribute to the originality of the approach. Indeed we propose a quite novel solution with a network of fixed sensors to collect audio signals (bird songs) that are interpreted in terms of individual positions in space and in time.

A. *Sensor network configuration*

In order to configure the wireless sensors network fitted with microphone, we assume that the target area is divided into sub-areas, that we call "regions" and that are supposed to be delimited by experts. Typically, a region may delimit an area around a water supply point. Another region may define an area of dense vegetation, or freshness, etc. The advantage of such a definition is that it eventually allows to obtain semantic information on the movements of individuals based on characteristics of regions. Unfortunately, this form of delimitation is not always possible. In a first stage, we define regions in an arbitrary manner.

Thus, each region has a unique identifier r_1, r_2, r_3, etc. Each sensor is attached to one region only. We will see that depending on the sensor placement method, a sensor can be in a region boundary and even in some cases, its detection area may overlap several regions. All sensors of a same region can communicate with each other through intra-region communications. For example, these communications are used to determine the number of birds in a region. The detail of methods that we defined for counting birds with a similar sensor network [11], [12], [24] are out of scope in this paper. They use triangulation techniques to locate songs and signal comparisons. Thus in this paper, we are able to consider that birds in a region may be counted on the basis of their songs.

Each region has at least one specific sensor called "gateway-sensor". It is able to communicate with other gateway-sensors of neighbor regions through inter-regions communications. These communications allow regions to exchange information and communicate with the central base responsible for collecting the whole records for further analysis. Figure 1 presents this architecture.

These two-layer communication system, intra-region and inter-region, not only enables sensors to locally determine the detected birds, but prevents that redundant information was

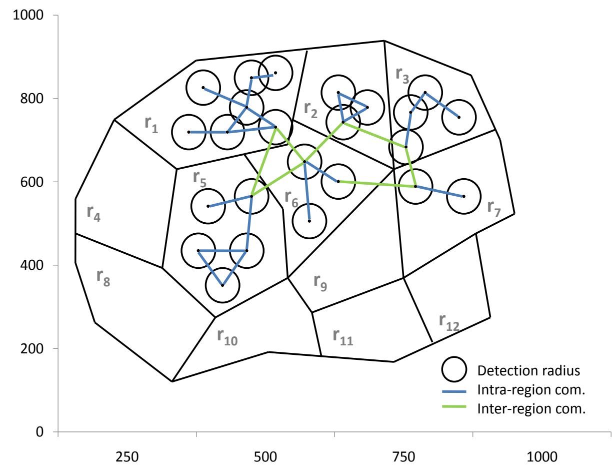

Fig. 1. Sensor architecture

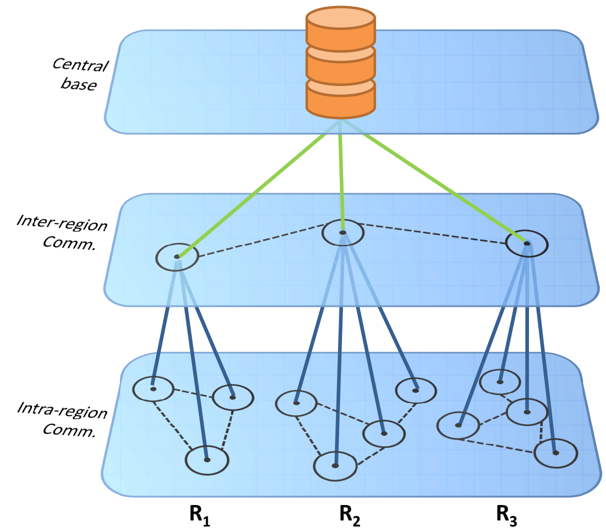

Fig. 2. Communication levels

sent to the central base. This has the effect of severely limiting useless exchanges in the network.

The two communication levels are presented on Figure 2. We can see an architecture on three stages :

1) The sensors communicate with the gateway-sensors.
2) The gateway-sensors communicate with the base station.
3) The base station stores aggregated data.

Figure 2 shows two types of communications. Colored links illustrate the communications at a logical level. Indeed, all detection sensors can send information to their gateway-sensors. Similarly, only the gateway-sensor can communicate with the central base.

In practice, a given sensor may not be able to communicate directly with its gateway-sensor. For instance, a sensor that is too far from its gateway-sensor of its region will have to contact a neighbor sensor to send information. These communications are real. They are represented by dotted lines and exist both between sensors and between gateway-sensors.

We assume that all sensors have the same properties: both same detection radius r and same energy level. Each sensor knows the region to which it is affiliated and its position in this region. Sensors are able to detect an individual of the given species by analyzing its song [12], [24], but they are not able currently to differentiate individuals between them. They can communicate with the sensors in the same region to send information to gateway-sensors. The gateway-sensor determines the number of birds currently in its region and sends this information to the base station. The central base knows all regions and receives the information on detected birds count in each region from gateway-sensors. In this work, we are not interested in the routing information to the base station. We focus only on the study of movements.

Before presenting the problem more formally, it is important to detail sensor placement, i.e. how to locate sensors against each other. Indeed, while only one region is assigned to a sensor, the detection radius of a sensor can cover several regions and thus strongly influences the results. There are different kinds of configuration:

Uniform configuration: this configuration consists on laying out the sensors uniformly so that they cover the entire zone. Unfortunately, this scheme is not realistic, in the case of dropping sensors by plane, as it is often the case.

Dense configuration: this configuration consists on placing a maximum of sensors to cover the maximum area. This configuration naturally raises problems of cost and interference between signals.

Incremental configuration: this method consists on optimizing the overall calculation, with an approximate optimal place for each new sensor integrated into the network. This mechanism is good for adjustment after installation.

Random configuration: this method consists on placing the sensors randomly on the area, usually by dropping of the sensors by plane.

Self-deployment: the objective of this method is to maximize the surface covered by the network, and ensure the inclusion of obstacles. The disposition by self-deployment should be done independently by the nodes.

Due to the specific constraints we meet, we are only able to consider a random deployment. However, we show below how the choice of a placement may influence the detection of movements.

B. Motion Detection Problem

The technical solution we adopted due to domain constraints provides only a limited set of collected information: a sound, the region where it has been recorded, an estimate of its position (using triangulation scheme) and the recording time. In this configuration, we assume that such information is collected over a period long enough to be exploited. Thus, whenever a desired sound is detected, for example the specific song of bird specie, communications intra and inter regions are done to send to the central base the information which will be used to study the movements. Then the base station stores a dataset which is analyzed and investigated to discover patterns in movements.

More formally, let denote $R = \{r_1, ..., r_n\}$ the set of all the regions which divide the area. Each gateway-sensor is able to count locally the number of birds at a time t_i on its region [11], [12], [24]. Thus, for a given time sequence $I = \langle t_1, ..., t_k \rangle$, the central base stores a $n \times k$ matrix M, called "detection matrix", in which each element M_{ij} represents the count in the region r_i at a time t_j.

Different patterns may be elicited from lines and columns of M: flock and periodic patterns. These data aggregated by regions, enables to handle the "noise" which may exist when considering individual positions. Indeed, the entities surveyed never have exactly the same coordinates while moving. Moreover, sensors that we use can only provide an estimate of the position by triangulation techniques. However, this region oriented approach is highly dependent on the size of regions. If an area is too large, the movement may be detected only in this region. On the other hand, if regions are too narrow, movements may occur in different regions. Anyway in the following subsections, we show how it is possible to find the two kinds of pattern with this sensor architecture.

IV. Flock patterns

Several definitions have been given for the expression of Flock pattern. Let us consider a sequence of successive times $I = \langle t_1, ..., t_k \rangle$. Benkert et al. [3], define a Flock pattern as being a movement of at least m entities such as for each time in I, we can find a disk of radius v which contains m entities. However, in the context of this paper, we have to take a less precise definition for two reasons. First, the current network does not supply the birds position with precision. We cannot use disks of a given radius to seek such groups. In addition, the detection is highly dependent on bird song and we know that at a time t_j, all individuals in the group does not necessarily sing. Thus, we are unlikely to find a group of the same size exactly at each point in I. That is why, in our context, we give another definition.

Definition 1: Let $\alpha \in \mathbb{N}$ be the detection margin of a group, $m \in \mathbb{N}$, I a time sequence $\langle t_1, t_2, ..., t_k \rangle$ with $\forall j \in [1..k]$, $t_j < t_{j+1}$, R the set of given regions and E is a set of m entities. We call $\alpha-$Flock pattern of m individuals over E, R and I, the sequence $\langle r_{t_1}, r_{t_2}, ..., r_{t_k} \rangle$ of regions in R such as at each time t_j of I, a maximal set of g entities among E occur in region r_{t_j} with $|m - g| \leq \alpha$.

Thus, we consider a $\alpha-$Flock pattern as a motion of almost the entire group observed. If the whole area is divided in regions so that they are not too large, we can consider that the birds appearing in same region at same time all along the I time sequence represent a flock movement. The example of the matrix M_1 below, shows the flock movement of six birds whose songs have been partially recorded since they are not singing at every time in I. Indeed, it is expected that the number of birds detected is not always equal to the group size. This is the result of two phenomena inherent in our solution based on sound detection. First, sensors are only fitted for sound recordings, but at a time t_j, birds in the group are unlikely singing together. Secondly, the noise that can occur on the signal may prevent the detection.

$$M_1 = \begin{bmatrix} 6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 5 & 6 \\ 0 & 0 & 5 & 5 & 0 & 5 & 5 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 6 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

According to the definition, an $\alpha-$Flock pattern may be seen as the sequence $S = \langle r_{t_1}, r_{t_2}, ..., r_{t_k} \rangle$ of regions, if it exists, that maximizes the number of entities detected. Thus $\forall r_{t_j} \in S$, r_{t_j} is the only region r_i of R, if it exists, such that $M_{ij} = \max_{l=1}^{n} M_{lj}$ and $|m - M_{ij}| \leq \alpha$. Thus, for this example if we take $\alpha = 2$, we detect a Flock pattern of a six birds across regions $\langle r_1, r_1, r_3, r_3, r_4, r_3, r_3, r_3, r_2, r_2 \rangle$.

For now, our architecture only allows us to study the collective movements of the entire group. According to the above definition, we consider that if at the same time t_j, several regions satisfy this definition, birds do not adopt a flock movement. Let us consider the example of matrix $M_1^{'}$ below.

$$M_1^{'} = \begin{bmatrix} 6 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 5 & 6 \\ 0 & 0 & 5 & 3 & 0 & 5 & 3 & 4 & 0 & 0 \\ 0 & 0 & 1 & 3 & 6 & 0 & 3 & 0 & 0 & 0 \end{bmatrix}$$

We can observe at time t_4 and t_7, no region satisfies the definition whatever be α. This may be due to three phenomena. Either the bird community does not adopt a flock movement, or the majority of birds does not sing enough or sensors do not correctly detect the location of birds in the regions. In all cases for such a recording, we consider that the birds do not follow a flock movement.

V. Periodic patterns

A periodic movement is a movement which repeats itself after a certain time called the period. A periodic pattern is not a flock pattern, it only reveals tendencies to adopt periodic paths through regions.

Definition 2: Let us consider S as a sequence of regions. We call periodic segment of period T, a sub-sequence $\langle r_{t_a}, ..., r_{t_b} \rangle$ of S, where $(t_b - t_a)$ *mod* $T = 0$ and for each $(t_p, t_q) \in \mathbb{N} \times \mathbb{N}$, $a \leq p \leq b$, $a \leq q \leq b$ we have $r_{t_p} = r_{t_q}$ if t_p *mod* T = t_q *mod* T. Thus the segment $s = \langle r_a, r_{a+1}, ..., r_{a+T} \rangle$ is repeated in S that is called a periodic pattern of period T.

The detection matrix M_2 below shows for instance a Periodic pattern observed for six birds even if the number of birds detected is not always equal to the size of the group.

$$M_2 = \begin{bmatrix} 4 & 6 & 0 & 0 & 0 & 6 & 6 & 0 & 0 & 0 \\ 1 & 0 & 5 & 1 & 0 & 0 & 0 & 6 & 1 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

The analysis of M_2 is done in two steps. As described for Flock pattern, we begin by extracting from M_2 the sequence S which maximizes the number of detected entities as follows: $\forall r_{t_j} \in S$, r_{t_j} is the region r_i of R such that $M_{ij} = \max_{l=1}^{n} M_{lj}$. Then, we extract from S the periodic segment s which repeats itself in S.

For instance, $S = \langle r_1, r_1, r_2, r_3, r_4, r_1, r_1, r_2, r_3, r_4 \rangle$, is deduced from M_2. The objective is to extract from S the periodic segment $\langle r_1, r_1, r_2, r_3, r_4 \rangle$. How can we find this segment ? Some algorithms had already been developed to mining periodic pattern in spatio-temporal data. We detail below the reason why these algorithms cannot be applied directly and we present a simple algorithm to discover periodic segment in S.

The example of matrix M_2 is very interesting. Indeed, if we define $\alpha = 2$, we can observe that the movement is both periodic and flock. But it is not always the case. Indeed, although our architecture only allows us to study collective movements, bird movements can be detected as being periodic type without be flock type. If we define $M_{2_{(5,4)}} = 1$ our specific definition allows us to detect a periodic pattern without detect a flock pattern.

This issue can be considered as apparently similar to questions studied in data stream or spatio-temporal data mining. The search for sequential patterns among large data streams has been conducted since earliest data mining works as an extension of classical association mining [1], [17], [28]. These patterns provide useful information on sequential relationships between objects in a dataset in multiple areas like web design, DNA sequences, intrusion detection or sensor network analysis for instance. In on line data streams, due to the continuous and high speed data flow it is not possible to store the whole data and to perform multi-scan of the dataset like in traditional data mining solutions. It is thus necessary to keep only the least possible data and to mine approximate patterns in order to avoid memory overflow [22]. Pattern mining in temporal and spatio-temporal data [14], [26], [21], [13], [15] is mainly based on time-series analysis where data are collection of time-series of each object over time. Indyk et al.[15] present this problem as follows. Given a long sequence S and a period T, the aim is to discover the most representative trend that repeat itself in S every T timestamps. Wang et al. [26] introduced flow patterns which describe the changes of events over space and time. They consider events occurring in regions and dependencies among changes in neighboring regions. The work conducted by Mamoulis et al. [21] on the search for periodic patterns focused on the movement of entities by using an area divided into regions too. They partition the space into a set of regions which allows them to define a pattern P as a sequence $[r_0, r_1, ..., r_n]$ of given length n, where r_i is a spatial region or the special character *, indicating the whole spatial universe. If the entities follow the pattern P enough times, the pattern is said to be frequent.

While this topic seems quite similar to the issues addressed in our case, the solutions are not suited to it for two reasons. The first one is that we do not have any information about the period of repetition of the pattern like in [21] or [6]. It is in fact what we want to determine. The second is that unlike the proposed methods, we do not have personal data on the movements of each individual, but we have only one estimate sequence of collective movements of birds. Indeed, the detection matrix M introduced in Section 3 only allows us to obtain the most representative sequence in movements. That's why we do not seek to obtain a frequent periodic segment for each individual, but rather a fixed segment, which is repeated in a collective sequence S. Indeed, if the group of birds really adopt a movement of periodic nature, we expect to identify a corresponding sub-segment into S that should be exactly the same at every period T.

This specific problem seems to have been thoroughly studied in bioinformatics [10], [6]. Indeed, the search for repeated patterns in DNA and protein sequences is proved to be important in sequence analysis, in particular in large-scale genome sequencing projects. Although these methods are efficient, they often aim to seek the repetition of segments known in advance or unequally spaced in the sequence.

We propose the simple algorithm ***Periodic Mine*** very much inspired by these methods, to dynamically generate a candidate segment from a sequence S and automatically check if it is periodic in S. This algorithm extracts a periodic segment $Scand$ from a sequence S. The rehearsal period T of the segment $Scand$ in S is equal to the size of the segment, i.e. $T = |Scand|$.

Input : A sequence S
Output : A periodic sequence $Scand$ extracted from S

```
Function PeriodicMine( S : Sequence) : Sequence
    isFrequent : boolean
    isFrequent ← false
    SCand : Sequence
    SCand ← ∅
    While (S ≠ ∅ and isFrequent = false) do
        add S[0] to Scand
        removeElement 0 from S
        isFrequent ← is(Scand periodic in S)
    done
    If (isFrequent) then
        return Scand
    end If
    return ∅
End
```

Algorithm 1: PeriodicMine for Mining periodic sequence

VI. Experimental evaluation

In order to understand the collective behavior of the Moqueurs Gorge Blanche which motivated this work, we search for useful information on their movements. This knowledge may be significant to reintroduce them in environments that

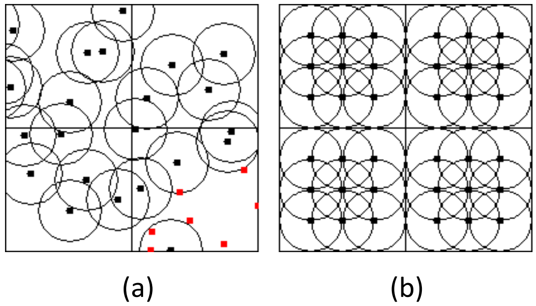

Fig. 3. random(a) and uniform(b) sensor placement in LYPUS

have similar characteristics to their original habitat. However, before deploying our solution on the ground in a real situation, a first necessary step is to design a prototype and thus measure the impact of all factors implicated on results. Indeed, to the best of our knowledge, the solution we propose has no equivalent in the literature. Consequently, simulations are expected to reveal relevant information for optimizing the deployment of wireless sensors. Simulations have indeed become essential steps in the study and comprehension of complex phenomena. They will allow us to test different configurations, but also to analyze parameters that are most prominent for the detection of movements. Thereafter, we plan to use the results obtained in simulation to adapt the architecture in real environment. This section is devoted to the simulation tool LYPUS and the results.

In order to evaluate the efficiency and the role of various parameters in the discovery of patterns, we built the 2D simulation environment LYPUS, which models the behavior of birds in their habitat. This simulation aims at representing a natural environment that virtually reproduces bird habitat in which virtual sensor network is set up.

In this tool, the area is first divided into uniform regions. Then, the sensors are set up and associated to the region to which it belongs. All sensors have the same capabilities (detection radius and battery level). We are interested nor in optimizing exchanges, neither in reducing loss or corruption of information frequently observed on sensor networks. We assume that inter and intra regions communications are always successful.

LYPUS is fully customizable. Thus, it is possible to define the size of the area, the number of regions and their size, the number of sensors, the size of their detection radius, the kind of sensor configuration, the size of the bird population, the probability of bird song, the type of movements, etc. Figure 3 shows the LYPUS 2D interface visualization with two areas of 200m x 200m. The largest rectangle represents the global area and each small rectangle represents a region. Sensors are figured by a point and their detection radius by circles. Birds are represented by small red filled squares.

To evaluate the efficiency of the architecture in pattern discovery, we give predefined behaviors to birds. We have implemented the two types of movements formalized in section 3: Flock and Periodic movements. The objective is to check and evaluate how the method we propose can retrieve these behaviors. Results which are proposed have been made with a calibration that we consider close to the real environment of birds. The dimension of the area is about 1000m x 1000m and it is divided in 100 regions of 100m x 100m. The sensors have detection radius of 25m. Furthermore, we assume that when several birds are singing in a sensor neighboring, it generates noise that cancels the recognition process. These birds are not detected by this sensor.

The results were obtained on a Intel Core 2 Duo P8600 2.4 Ghz, with 4Go RAM, Linux Ubuntu 9.04 and JDK 1.6. The pattern discovery efficiency is evaluated from the average results of 100 tests, using different configurations for sensor location and route taken by birds. The discovery of the $\alpha-$Flock pattern is done with $\alpha = 2$.

However, it is clear that the discovery of patterns with such an architecture is confronted with three main issues: the song occurrence, the coverage of the area and the kind of placement. It is obvious that if birds do not sing, we will not detect any pattern. Moreover, according to the way sensors are set up, either the area may be partially covered or sensor detection regions may overlap. This situation may create a bias on the process of pattern discovery. That is why we have studied the impact of these three parameters.

First, performances were evaluated with random configuration (Figure 3(a)) for different network size as shown in Figure 4. Then we analyzed the impact of the population size on both random and uniform (Figure 3(b)) configuration as show in Figure 5.

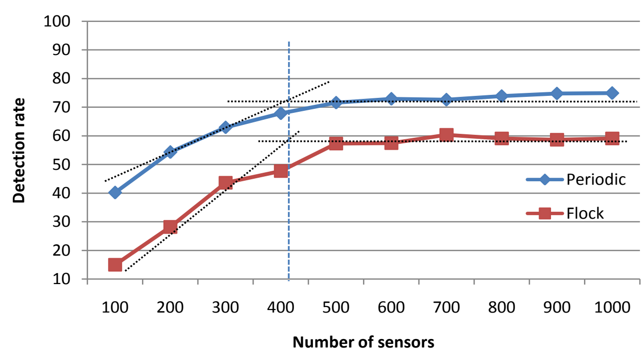

Fig. 4. Detection rate with different network size

Figure 4 shows the correct detection rate of the two patterns when the sensors number varies with a population of 20 individuals. We notice that when the number of sensors is increasing, the detection rate is improving too. This is because the area is rather wide and when the sensors number increases, the surface coverage becomes more important and the detection of the two patterns is improved. From a certain threshold, we can observe a plateau for which improved detection becomes insignificant. We also notice that the results are better for Periodic pattern than for $\alpha-$Flock pattern.

We are aware that a major issue related to such an architecture is the probability of song occurrence. Indeed, the

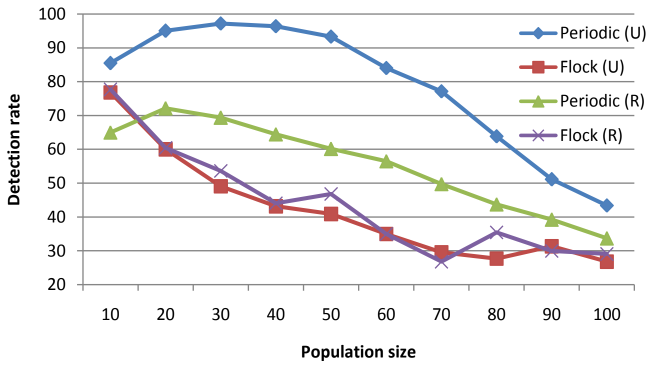

Fig. 5. The impact of population size and the kind of placement

song and its location are the only data that sensors can currently collect. However, we know that birds do not sing all the time. To study the impact of this phenomenon, we have chosen to make the population size vary, rather than the probability of singing. Thus if we increase the size of the population, we increase the probability that several individuals sing. These results are shown in Figure 5 for an area of 1000m x 1000m divided in 100 regions where two configurations of 900 sensors are compared: random and uniform sensors layer. Indeed, the assumption of random placement of sensors is the most realistic, but this method produces unbalanced sensor distribution over regions and some sensors may overlap several regions (Figure 3).

On Figure 5 **Flock(R)** and **Periodic(R)** are the results obtained for a random configuration and **Flock(U)** and **Periodic(U)** are the results obtained for a uniform configuration.

Figure 5 shows that for both configurations, when the bird population increases, the detection rate decreases. Indeed, when the number of birds is increasing, the probability that several birds sing at the same time on the same region is more important. This phenomenon generates noise in the sensors neighboring. In [24], we showed that this problem could be partly solved by adding sensors.

Although a uniform placement is not conceivable in practice, we can compare these results with those obtained in the same conditions for the random placement. First, we can notice that the Flock pattern detection is very close for the two configurations. However, Periodic Pattern detection is improved. Indeed, we obtain rates higher than 90 % for populations under 50 individuals.

In conclusion, these results highlight the dependency of the proposed method efficiency on these different factors: the region size, the probability of singing, the population size and the number of sensors. However, these results seem to show that an optimal configuration must be studied as a function of the population size. The population size appears to be the most relevant factor that influences results. So it would be possible, if the population size is known or approximated, to adjust the placement of sensors in order to achieve optimal detection of bird movements.

VII. Conclusion and Future Work

In this work, we have presented a new framework based on sensors fitted with microphone for collecting data on collective bird movements. Unlike methods that use devices grafted to each animal, our solution not only preserve bird communities but is able to provide relevant information on collective behavior too. Our contribution can be summarized as follows :

- We have discussed the problem of studying collective movement patterns and the main techniques currently used.
- Then we have proposed an architecture of sensors fitted with microphone to collect data on bird movements. We have presented the challenges related to such an architecture and solutions to optimize exchanges and processing of data on the network.
- Thereafter we have shown how collected data can be exploited to search for patterns in birds movement.
- Finally, we have studied experimental results obtained by simulation and the impact of different parameters on the results.

This work is a preliminary and mandatory step in this project which will consist on studying and modeling animal behavior, by means of wireless sensor technologies. In the short term we plan to adapt the framework to the real environment. In the long term we want to improve the detection of movement by fitting sensors with other devices such as cameras.

Acknowledgements

We thank J.-R. Gros-Desormeaux from the GEODE[1] laboratory for motivating discussions. He gave us pertinent information about birds and their behaviors.

References

[1] R. Agrawal and R. Srikant. Mining sequential patterns. *Proceeding of the 11th International Conference on Data Engineering, Taipei, Taiwan*, pages 3–14, 1995.

[2] R.E. Bell. A sound triangulation method for counting barred owls. *The Wilson Bulletin. The Wilson Ornithological Society*, 1964.

[3] M. Benkert, J. Gudmundsson, F. Hubner, and T. Wolle. Reporting flock pattern. *Computational Geometry: Theory and Applications*, pages 111–125, 2008.

[4] L. Brotons and S. Herrando. Effect of increased food abundance near forest edges on flocking patterns of coal tit parus ater winter groups in mountain coniferous forests: Individuals concentrated near forest edges in bigger social groups than in forest interiors and foraged more on pine cones which were more abundant there. *Bird Study*, 50:106–111, 2003.

[5] H. Chok and L. Gruenwald. Spatio-temporal association rule mining framework for real-time sensor network applications. *Conference on Information and knowledge management*, pages 1761–1764, 2009.

[6] E. Coward and F. Drablos. Detecting periodic patterns in biological sequences. *Bioinformatics*, 14:498–507, 1998.

[7] A. Desrochers and M.J. Fortin. Understanding avian responses to forest boundaries: a case study with chickadee winter flocks. *Oikos*, 91:376–384, 2000.

[1] GEOgraphie - Developpement - Environnement, University of Antilles and Guyane

[8] A. Dolby. Movement patterns of male common voles (microtus arvalis) in a network of y junctions: role of distant visual cues and scent marks. *National Research Council of Canada, Ottawa*, 79(12):2228–2238, 2001.
[9] B. Dumont, A. Boissy, C. Achard, A. Silbbald, and H. W. Erhard. Consistency of animal order in spontaneous group movements allows the measurement of leadreship in a group of grazing heifers. *Applied animal behavior science*, 95(1-2):55–66, 2005.
[10] E. F. Glynn, J. Chen, and A. R. Mushegian. Detecting periodic patterns in unevenly spaced gene expression time series using lomb scargle periodograms. *Bioinformatics*, 22(3):310–6, 2006.
[11] H. Gros-desormeaux, P. Hunel, and N. Vidot. Counting birds with wireless sensors networks. *IWCMC 2009 Wireless Sensor Networks Symposium*, pages 1163–1167, 2009.
[12] H. Gros-desormeaux, P. Hunel, N. Vidot, and E. Stattner. Acoustic counting algorithms for wireless sensors networks. *Proceedings of the 6th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks*, pages 79–84, 2009.
[13] J. Gudmundsson, P. Laube, and T. Wolle. Movement patterns in spatio-temporal data. *Encyclopedia of GIS*, pages 726–732, 2008.
[14] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. *International Conference on Data Engineering*, pages 106–115, 199.
[15] P. Indyk, N. Koudas, and S. Muthukrishan. Identifying representative trends in massive time series data sets using sketches. *Proc. of Very Large Data Base*, pages 363–372, 2000.
[16] C. S. Jensen, D. Lin, and B. C. Ooi. Continuous clustering of moving objects. *Knowledge and Data Engineering, IEEE Transactions on*, 19:1161–1174, 2007.
[17] N. Jiang and L. Gruenwald. Research issues in data stream association rule mining. *ACM SIGMOD Record*, 35:14–19, 2006.
[18] P. Laube and S. Imfeld. Analyzing relative motion within groups of trackable moving point objects. *Proceedings of the Second International Conference on Geographic Information Sciences*, pages 132–144, 2002.
[19] P. Laube, M. van Kreveld, and S. Imfeld. Detecting relative motion patterns in geospatial lifelines. *International Symposium on Spatial Data Handling*, pages 201–214, 2004.
[20] Y. Li, J. Han, and J. Yang. Clustering moving objects. *International Conference on Knowledge Discovery and Data Mining*, pages 617–622, 2004.
[21] N. Mamoulis and M. Hadjieleftheriou. Mining, indexing, and querying historical spatiotemporal data. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '04*, pages 236–245, 2004.
[22] A. Marascu and F. Masseglia. Mining sequential patterns from temporal streaming data. *Proceedings of the first ECML/PKDD Workshop on Mining Spatio-Temporal Data (MSTD '05), held in conjunction with the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD '05*, 2005.
[23] M. A. Rumble, L. Benkobi, F. Lindzey, and R. Scott Gamo. Evaluating elk habitat interactions with gps collars. *International conference held at the Macaulay Land Use Research Institute*, pages 11–17, 2001.
[24] E. Stattner, P. Hunel, N. Vidot, and M. Collard. Efficient scheme to count bird song with wireless sensors networks. *Internal report: University of Antilles and Guyane, LAMIA*, pages 1–9, 2010.
[25] F. Verhein and S. Chawla. Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases. *Proceedings of the 11th International Conference on Database Systems for Advanced Applications*, 3882:187–201, 2006.
[26] H. Wang, W. Fan, P. S. Yu, and J. Han. Mining concept driftingdata streams using ensemble classifiers. *International Conference on Knowledge Discovery and Data Mining*, pages 226–235, 2003.
[27] J. C. Xia, V. Ciesielski, and C. Arrowsmith. Data mining of tourists' spatio-temporal movement patterns —a case study on philip island. *Proceedings of the Eighth International Conference on GeoComputation*, pages 1–15, 2005.
[28] Y. Zhu and D. Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. *Int'l Conf. on Very Large Data Bases*, 2002.