



Co-movement Pattern Mining from Videos

Dongxiang Zhang
The State Key Laboratory of
Blockchain and Data Security,
Zhejiang University
zhangdongxiang@zju.edu.cn

Teng Ma
Zhejiang University, China
mt0228@zju.edu.cn

Junnan Hu
Zhejiang University, China
jnhu@zju.edu.cn

Yijun Bei*
Zhejiang University, China
beiyj@zju.edu.cn

Kian-Lee Tan
National University of Singapore
tankl@comp.nus.edu.sg

Gang Chen
Zhejiang University, China
cg@zju.edu.cn

ABSTRACT

Co-movement pattern mining from GPS trajectories has been an intriguing subject in spatial-temporal data mining. In this paper, we extend this research line by migrating the data source from GPS sensors to surveillance cameras, and presenting the first investigation into co-movement pattern mining from videos. We formulate the new problem, re-define the spatial-temporal proximity constraints from cameras deployed in a road network, and theoretically prove its hardness. Due to the lack of readily applicable solutions, we adapt existing techniques and propose two competitive baselines using Apriori-based enumerator and CMC algorithm, respectively.

As the principal technical contributions, we introduce a novel index called temporal-cluster suffix tree (TCS-tree), which performs two-level temporal clustering within each camera and constructs a suffix tree from the resulting clusters. Moreover, we present a sequence-ahead pruning framework based on TCS-tree, which enables the concurrent utilization of all pattern constraints to filter candidate paths. Finally, to reduce verification cost on the candidate paths, we propose a sliding-window based co-movement pattern enumeration strategy and a hashing-based dominance eliminator, both of which are effective in avoiding redundant operations.

We conduct extensive experiments for scalability and effectiveness analysis. Our results validate the efficiency of the proposed index and mining algorithm, which runs remarkably faster than the two baseline methods. Additionally, we construct a video database with 1169 cameras and perform an end-to-end pipeline analysis to study the performance gap between GPS-driven and video-driven methods. Our results demonstrate that the derived patterns from the video-driven approach are similar to those derived from groundtruth trajectories, providing evidence of its effectiveness.

PVLDB Reference Format:

Dongxiang Zhang, Teng Ma, Junnan Hu, Yijun Bei, Kian-Lee Tan, and Gang Chen. Co-movement Pattern Mining from Videos. PVLDB, 17(3): 604 - 616, 2023.
doi:10.14778/3632093.3632119

*Corresponding author.

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

Proceedings of the VLDB Endowment, Vol. 17, No. 3 ISSN 2150-8097.
doi:10.14778/3632093.3632119

PVLDB Artifact Availability:

The source code, data, and/or other artifacts have been made available at <https://github.com/Mateng0228/Co-movement-Pattern-Mining-from-Videos>.

1 INTRODUCTION

As a prevalent analytical challenge in spatial-temporal data mining, the discovery of co-movement patterns involves identifying all groups of objects that move within spatial proximity for a specified time interval. Several co-movement patterns have been proposed, each with distinct constraints in terms of spatial proximity and temporal consecutiveness. For instance, the flock [8, 24] and group [25] patterns require all objects in a group to be enclosed by a disk with a radius r . On the other hand, the convoy [12, 16, 17], swarm [14], and platoon [13] patterns adopt a more relaxed density-based spatial clustering approach to determine spatial proximity. Regarding the temporal dimension, flock and convoy patterns require global consecutiveness, where all timestamps of candidate groups must be consecutive. In contrast, group and platoon patterns tolerate local gaps between consecutive segments.

This co-movement pattern mining problem has found practical applications in animal behavior studies such as bird migration monitoring [24]. In addition, government agencies can leverage the co-movement pattern mining task to support smart city management, such as detecting traffic congestion at varying levels of granularity [17], computing evacuation schedules during disaster [2], security surveillance upon suspicious groups [26], and assessing the viability of launching public transport services in areas with dense co-movement patterns [16]. Prior works in this domain rely on large-scale GPS trajectories, which are primarily collected and owned by commercial hail-riding or map-service companies. However, in practice, these GPS-based trajectory data are not directly accessible to government agencies. Therefore, in this paper, we migrate the data source from GPS sensors to surveillance cameras and present the first work on co-movement pattern mining against an urban-scale video database. The idea is feasible, because with the rapid development of smart cities, numerous surveillance cameras have been deployed throughout the road network. These infrastructures and the video big data they generate are owned by government agencies. Given the availability of computation-efficient and accurate visual inference models, these data sources can be leveraged to support the aforementioned applications of co-movement pattern mining.

Given a corpus of video data captured by the surveillance camera network, we can extract the movement paths of individual objects using trajectory recovery algorithms [15, 21, 27]. In contrast to GPS trajectories, the extracted path is represented by a sequence of camera identifiers and time intervals $(c_1, [t_1^s, t_1^e]) \rightarrow (c_2, [t_2^s, t_2^e]) \rightarrow \dots \rightarrow (c_m, [t_m^s, t_m^e])$, where c_i denotes the camera capturing the moving object and $[t_i^s, t_i^e]$ represents the period during which the object is detected by c_i . In this study, we assume that the trajectory extraction has been performed by a black-box algorithm and concentrate on devising efficient algorithms for mining co-movement patterns. Notably, the practice of treating video ingestion as a pre-processing step has also been adopted by recent video mining studies [4, 5].

In our scenario, the spatial location information outside the intervals of $[t_i^s, t_i^e]$ is not available. This property hinders the application of original trajectory-based co-movement patterns and calls for problem reformulation. To determine spatial proximity from a video database, we leverage temporal proximity at the same camera as an approximation. In other words, if two moving objects are captured by the same camera within time interval Δ_t , we consider them spatially close to each other. As to the temporal duration for spatial proximity, we replace it with a sequence of k cameras. In other words, we aim to identify platoons of objects that travel together along the same route in the road network.

Figure 1 illustrates a toy example. Three distinct objects, namely o_1 , o_2 , and o_3 , traverse different routes on the road network. Specifically, o_1 and o_2 follow the same trajectory, which is $c_1 \rightarrow c_2 \rightarrow c_4 \rightarrow c_5$, while o_3 travels from route $c_3 \rightarrow c_2 \rightarrow c_4 \rightarrow c_5$. These objects are captured by the cameras at different time intervals. Suppose we set the minimum group size as 2 and require that the minimum temporal gap at a camera to determine spatial proximity is 5, and the objects must pass at least 3 cameras together. Based on these conditions, we find that the temporal gap between o_1 and o_2 is too large to form a valid co-movement pattern even though they share the same travel route. The minimum gap between their time intervals at camera c_4 is 7. Conversely, o_2 and o_3 are spatially close to each other on the path $c_2 \rightarrow c_4 \rightarrow c_5$, and we can identify $\langle\{o_2, o_3\}, c_2 \rightarrow c_4 \rightarrow c_5\rangle$ as a valid co-movement pattern.

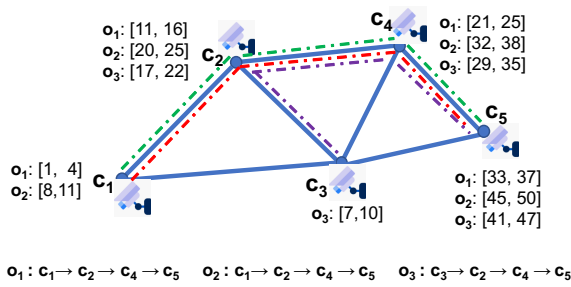


Figure 1: A toy example of co-movement pattern mining from surveillance cameras.

In this paper, we formulate the problem of co-movement pattern mining and formally define the spatial-temporal proximity constraints arising from video data. This mining problem is challenging because of the potentially enormous search space that results from

a large number of moving objects travelling over a prolonged period across a road network with a plethora of surveillance cameras. We establish the hardness of this problem by demonstrating that it is NP-Hard through the reduction of the maximum clique problem to our pattern mining problem.

Given the absence of readily applicable solutions, we draw upon the ideas from prior work and develop two baseline algorithms. The first algorithm is inspired by SPARE [6], a parallel and general co-movement pattern miner from trajectories. We adopt its Apriori-based enumerator to systematically explore the search space. The algorithm can be viewed as *camera-ahead* exploration that enumerates candidate paths with increasing length. The remaining pattern constraints are then leveraged for further pruning. The second baseline extends the Coherent Moving Cluster (CMC) algorithm proposed in [12] for convoy pattern mining, with two tailored improvements for the video-based setting. The algorithm can be viewed as *temporal-ahead* since it searches across the temporal dimension and dynamically updates the candidate patterns.

As the core technical contributions, we propose a new index called temporal-cluster suffix tree (TCS-tree), which performs two-level temporal clustering within each camera. We represent the trajectory of each moving object as a sequence of cluster IDs and build a suffix tree to facilitate frequent subsequence mining. Based on TCS-tree, we propose a sequence-ahead pruning framework that enables the concurrent utilization of all pattern constraints to filter candidate travel paths. Finally, to minimize the verification cost of candidate paths, we propose a sliding-window-based co-movement pattern enumeration strategy and a hashing-based dominance eliminator, both of which effectively eliminate redundant operations.

In summary, we make the following contributions:

- (1) This study pioneers the investigation of co-movement pattern mining from video databases. We formulate this new mining problem and theoretically prove its NP-hardness.
- (2) We propose a novel index called temporal-cluster suffix tree (TCS-tree), an efficient sequence-ahead pruning framework to facilitate pruning, as well as an effective verification scheme based on a sliding-window based candidate enumeration strategy and a hashing-based dominance eliminator.
- (3) We conduct extensive experiments to validate the superiority of our proposed algorithm over the baselines. We also construct a video database with 1169 cameras to perform an end-to-end pipeline analysis and study the performance gap between GPS-driven and video-driven co-movement pattern miners.
- (4) To benefit the research community, we make all datasets and implementation code available on github¹.

The remainder of this paper is organized as follows. We formulate the new mining problem and theoretically prove its hardness in Section 2. Section 3 provides a survey of related works on co-movement pattern mining. The two baseline algorithms are presented in Section 4 and we propose our index and mining framework in Section 5. In Section 6, we conduct comprehensive experiments for performance evaluation. Finally, we conclude the paper and discuss potential avenues for future research in Section 7.

¹<https://github.com/Mateng0228/Co-movement-Pattern-Mining-from-Videos>

2 PROBLEM DESCRIPTION

2.1 Data Model

In our model for video database, a multitude of surveillance cameras is deployed throughout the urban city. These cameras capture a vast amount of objects in motion over different time periods. For ease of presentation, we assume that the cameras are not overlapped and each moving person or vehicle can only be captured by one camera. The extension to support overlapped cameras will be discussed in Section 5.5. As the data source solely comprises camera meta-data (e.g., location, resolution, and view angle) and the recorded video content, GPS trajectories of moving objects are not available directly. Nonetheless, we can employ recent trajectory recovery algorithms [15, 21, 27] to extract the travel routes of moving objects across different cameras. As such, to support spatio-temporal pattern mining, we define the travel path of a moving object o_i in the form of a camera id sequence.

Definition 2.1. Travel Path

We define the travel path of an object o_i as a sequence of camera IDs, arranged in order of the time periods in which o_i was captured: $P_i = (c_1, [s_1, e_1]) \rightarrow (c_2, [s_2, e_2]) \rightarrow \dots \rightarrow (c_n, [s_n, e_n])$, where $s_j < e_j < s_{j+1}$ and $(c_j, [s_j, e_j])$ implies that o_i appears in c_j in the time interval $[s_j, e_j]$.

We call s_j the entrance timestamp and e_j the exit timestamp. The time interval of P_i is $[s_1, e_n]$. In cases without the need to mention the temporal dimension, we simplify the path representation as $P_i = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n$.

Definition 2.2. Sub-path

We say P_1 is a sub-path of P_2 , denoted by $P_1 \subseteq P_2$, if the camera sequence of P_1 is a subsequence of P_2 and the time interval of P_1 is contained by P_2 .

From the travel paths of all objects, we can construct a camera network in a data-driven manner. Its vertices contain all disjoint cameras that appear in the travel paths. Two cameras c_i and c_j are connected in the camera network as long as there exists a travel path containing $c_i \rightarrow c_j$. The data model of camera network will be used in the two baseline algorithms.

2.2 Mining Task Formulation

To define co-movement patterns over the travel paths of moving objects, we notice that the traditional definition of co-movement patterns involves a parameter ϵ to determine spatial proximity, where ϵ is either the radius of the disk or the threshold for density reachability. However, in the scenario considered in this paper, the exact GPS locations of moving objects are not available, and we must rely on other clues to determine spatial proximity. In the following definition, we utilize temporal proximity at the same camera to infer spatial proximity.

Definition 2.3. ϵ -reachability at camera c_j

Two objects o_1 and o_2 are ϵ -reachable at camera c_j if the gap between their entrance timestamp at c_j is no greater than ϵ .

The measurement of temporal duration, such as “ k consecutive timestamps” defined in convoy and flock patterns, is not applicable in our scenario because the temporal dimension of travel path P_i

only involves a sequence of disjoint time intervals. In our problem definition, we replace the constraint of temporal duration with “ k consecutive cameras in the travel paths” of the grouped objects.

Definition 2.4. Co-movement Pattern from Videos

Given parameters ϵ , m and k , a co-movement pattern CP in this paper is defined as $\langle O_i, P_i \rangle$, where O_i is an object set and P_i is a travel path. They satisfy the following constraints:

- (1) O_i contains at least m objects, i.e., $|O_i| \geq m$.
- (2) P_i contains at least k cameras, i.e., $|P_i| \geq k$.
- (3) For each object $o \in O_i$ with travel path P_o , P_i is a sub-path of P_o , i.e., $P_i \subseteq P_o$.
- (4) For any two objects $o_i, o_j \in O_i$ and camera $c_i \in P_i$, o_i and o_j are ϵ -reachable at camera c_i .

Under certain parameter settings, such as small values for m and k or a large value of ϵ , it is possible that an extensive number of valid patterns may arise. In this paper, we follow previous approach [17] to study maximal co-movement pattern mining and define pattern dominance as following.

Definition 2.5. Pattern Dominance

A co-movement pattern $CP_1 = \langle O_1, P_1 \rangle$ is dominated by $CP_2 = \langle O_2, P_2 \rangle$ if $O_1 \subseteq O_2$ and $P_1 \subseteq P_2$.

Our objective is to identify all the co-movement patterns $CP_i = \langle O_i, P_i \rangle$ that are both valid and non-dominated. The hardness of the problem is proved in the following theorem.

THEOREM 2.6. *The problem of maximal co-movement pattern mining in video database is NP-hard.*

PROOF. We prove the theorem by reducing the maximum clique problem to our pattern mining problem.

Let $G = (V, E)$ be a graph with n vertices. We conduct the reduction as follows. Each vertex $v_i \in G$ is uniquely mapped to a moving object o_i in our problem. Let $\eta = \max(|V|, k)$ and we make all moving objects pass through the same camera sequence $c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_\eta$. The entrance time of object o_i at camera c_j is defined as

$$t_{ij} = \begin{cases} (2\epsilon + 1) * j - \epsilon & i \neq j \wedge i \text{ and } j \text{ are not connected in } G \\ (2\epsilon + 1) * j & i \neq j \wedge i \text{ and } j \text{ are connected in } G \\ (2\epsilon + 1) * j + \epsilon & i = j \end{cases} \quad (1)$$

With the above settings, we know that if v_a and v_b are not connected in G , the entrance time of o_a at camera c_b is $(2\epsilon + 1) * b - \epsilon$ and the entrance time of o_b at camera c_b is $(2\epsilon + 1) * b + \epsilon$. These two objects are not ϵ -reachable at camera c_b and will not be co-located within a valid co-movement pattern. On the other hand, if v_a and v_b are connected in G , we show that $|t_{aj} - t_{bj}| \leq \epsilon$ for all cameras $c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_\eta$:

- (1) If $j = a$ or $j = b$, $|t_{aj} - t_{bj}| \leq |(2\epsilon + 1) * j + \epsilon - (2\epsilon + 1) * j| \leq \epsilon$.
- (2) If $j \neq a$ and $j \neq b$, $|t_{aj} - t_{bj}| \leq |(2\epsilon + 1) * j - \epsilon - (2\epsilon + 1) * j| \leq \epsilon$.

Therefore, if there exists a maximum clique with size m , we can also find a co-movement pattern covering m objects through η consecutive cameras. Obviously, this co-movement pattern is maximal. \square

2.3 End-to-End Mining Pipeline

The complete mining pipeline that starts from the input of surveillance videos includes three major steps:

Step 1: Pre-processing. The goal is to extract the cross-camera trajectories from the surveillance cameras deployed in a road network. This step can be achieved by adopting existing trajectory recovery algorithms from large-scale videos [15, 21, 27] or multi-camera multi-target tracking models [10, 11, 20]. The latter is more accurate but consumes higher computation cost. The trajectory extraction can be viewed as a pre-processing step and the extracted trajectories can be used to support co-movement pattern mining or video database queries.

Step 2: Online Index Construction. After extracting the camera sequences for all moving objects, we can build index in an online manner to facilitate co-movement pattern mining. For example, we will build a temporal-clustering suffix tree, which relies on the query threshold ϵ .

Step 3: Pattern Mining. Users can apply the mining algorithms proposed in this paper to retrieve all the co-movement patterns satisfying the spatial-temporal proximity constraints. For parameter setting, ϵ can be set from 1-4 minutes to take into account the effect of traffic lights. Parameters m and k can be determined by specific applications. Since the co-movement pattern mining is an exploratory process, if the query user is not satisfied with the results, he/she can adjust the parameters and proceed to step 2 for the next round of pattern mining.

For ease of reference, the notations frequently used in the paper are summarized in Table 1.

Table 1: Notation Table.

c_i, o_i	A surveillance camera c_i and a moving object o_i
P_i	The travel path of object o_i
O_i	An object set
ϵ	The threshold for temporal closeness
m	The minimum group size for co-movement pattern
k	The minimum route length for co-movement pattern
CP	A co-movement pattern
TC	A temporal cluster represented as triplet (O, l, r)
TC_{c_i}	The temporal clusters in camera c_i

3 RELATED WORK

Among the co-movement patterns defined over GPS trajectories, the flock [8, 24] and convoy [12, 16, 17] patterns require the candidate groups to appear in k consecutive timestamps. Their difference lies in the definition of spatial proximity. Flock requires objects in the same cluster to be within a disk-region of diameter less than parameter ϵ , while convoy uses density-based spatial clustering. In their mining algorithm designs, performing clustering at every timestamp is computationally expensive and can become a performance bottleneck. Several techniques have been proposed to alleviate this overhead, such as spatial partitioning [16, 24], trajectory simplification [12], and the divide-and-conquer approach [17].

Different from flock and convoy, the group [25], swarm [14] and platoon [13] patterns have more relaxed constraints on the pattern

duration. Their techniques of mining are of the same skeleton. The main idea of mining is to grow an object set from an empty set in a depth-first manner. Throughout this construction process, various pruning techniques are employed to discard unnecessary branches. Group pattern mining [25] uses a VGgraph to guide the pruning of false candidates, while swarm mining [14] relies on two more pruning rules called backward pruning and forward pruning. The platoon mining algorithm in [13] takes advantage of a prefix table structure to steer the depth-first search.

For co-moving pattern mining based on trajectory streaming, CoMing [3, 7] leverages Apache Flink for streaming-based processing, builds spatial indexes to accelerate clustering, and employs pruning techniques to eliminate unnecessary verification. In [19], the group pattern is relaxed to allow membership withdrawal or re-join as long as some participants stay connected for all time intervals. An incremental discovery solution is developed to retrieve the evolving companion efficiently from trajectory streaming data.

Recently, video-based pattern mining has emerged as an interesting database research topic. In [15, 21, 27], the task of snapshot clustering is studied to recover trajectories from urban-scale videos. The objective is to cluster snapshots of detected objects that are visually similar and spatially coherent. The items within each cluster are considered to belong to the same vehicle. They are sorted by the associated timestamp to form a trajectory. In [4, 5], Chen et al. retrieved video segments that satisfy user-provided constraints on the spatial or temporal relationships among objects of interest, as well as other conditions on the object labels.

The differences between GPS-based and video-based co-movement pattern mining are summarized as:

Table 2: Differences between GPS-based and video based co-movement pattern mining.

	Trajectory-based	Video-based
Element in trajectory	$\langle lat, lng, time \rangle$	$\langle c_j, [s_j, e_j] \rangle$
Trajectory representation	synchronized and fine-grained	asynchronized and coarse-grained
Proximity Determination	spatial clustering	temporal clustering within a camera
Duration Determination	consecutive timestamps	k consecutive paths in the camera network
Performance bottleneck	spatial clustering	candidate verification

4 TWO BASELINE ALGORITHMS

In this section, we present two baseline algorithms for co-movement pattern mining, by adapting the ideas from trajectory-based solutions to the context of video database.

4.1 CMC Algorithm

The Coherent Moving Cluster (CMC) algorithm [12, 18] represents a simple and general baseline technique for mining convoy patterns. The algorithm operates by performing spatial clustering on the moving objects at each timestamp, followed by a search for all valid patterns across the temporal dimension. To this end, a global buffer of candidate patterns is maintained. At each timestamp, each

cluster is intersected with the partial patterns in the global buffer, thus generating new patterns that adhere to the constraints of k consecutive timestamps and minimum group size.

In this paper, we present a competitive baseline, which builds upon CMC with two key improvements tailored for video-based settings. Firstly, we substitute the traditional spatial clustering approach with a temporal clustering approach that ensures objects within the same cluster are ϵ -reachable. The temporal clustering is performed independently within each camera. Secondly, we replace the global buffer with distributed buffers across the cameras. At each timestamp, when accessing a temporal cluster within camera c_i , we intersect it with only the patterns in the local buffer of c_i , thus reducing the search space and enhancing computational efficiency. The newly generated patterns are propagated to the buffers in the neighboring cameras for future exploration.

4.1.1 Temporal Clustering. Recall that we utilize ϵ -reachability in the temporal dimension as an approximate means of representing spatial proximity. For the set of objects captured by camera c_i , we propose a lightweight temporal clustering approach that ensures objects within the same cluster are ϵ -reachable. Additionally, its linear complexity can prevent the clustering process from becoming a performance bottleneck.

As demonstrated in Algorithm 1, we access the moving objects in ascending order of their entrance time into camera c_i , utilizing two anchors, l and r to establish the boundaries of each cluster. To store the resulting clusters of camera c_i , we maintain a data structure $\mathbb{TC}_{c_i} = \{TC_1, TC_2, \dots, TC_i, \dots\}$, where each temporal cluster $TC_i = (O_i, [l_i, r_i])$ is associated with an object set O_i and time interval $[l_i, r_i]$. By employing the anchor r to determine the right boundary, all moving objects within temporal distance ϵ are grouped into a single cluster. We exclude small clusters that contain fewer than m objects.

Upon examining the temporal clustering, we can draw three observations. First, as each object is accessed by the anchor points at most twice, the time complexity of temporal clustering is linear to the number of objects captured by camera c_i . Assuming there are N objects, each with an average path length of L , the complexity of temporal clustering on all cameras is $O(NL)$. Second, any two objects within the same temporal cluster are guaranteed to be ϵ -reachable, as specified by line 5 in Algorithm 1. Finally, these temporal clusters are not necessarily disjoint in the temporal dimension. Thus, it is possible that an object in a camera belongs to multiple temporal clusters.

4.1.2 Candidate Pattern Enumeration. The search algorithm explores the temporal dimension for candidate pattern enumeration. We sort the temporal clusters $TC_i = (O_i, [l_i, r_i])$ from all cameras in ascending order of their starting time l_i . Let V_{in}^i denote the incoming neighbors of camera c_i in the camera network and V_{out}^i denote its outgoing neighbors. We maintain a local buffer for each camera to store candidate patterns whose path ends at any node in V_{in}^i . Suppose we are presently examining the temporal cluster $TC_i = (O_i, [l_i, r_i])$ derived from camera c_u , we perform object intersection to generate new candidate patterns. This step is similar to the original CMC algorithm which expands the time interval of

Algorithm 1: Temporal Clustering

```

1  $\mathbb{TC}_{c_i} \leftarrow \emptyset; TC \leftarrow (\emptyset, [0, 0]); r \leftarrow -1; l \leftarrow -1;$ 
2  $S \leftarrow$  object records in order of entrance time in camera  $c_i$ ;
3 while  $r \leq |S|$  do
4   if  $S[r].time - S[l].time \leq \epsilon$  then
5     Enqueue  $S[r].obj$  into  $TC.O$ ;
6      $TC.r \leftarrow S[r].time; r++$ ;
7   else
8     if  $|TC| \geq m$  then
9        $\mathbb{TC}_{c_i} \leftarrow \mathbb{TC}_{c_i} \cup TC$ 
10    while  $l \neq r$  do
11      if  $S[r].time - S[l].time \leq \epsilon$  then
12        break;
13      Dequeue  $S[l].obj$  from  $TC.O$ ;
14       $l++$ ;  $TC.l \leftarrow S[l].time$ 
15 if  $|TC| \geq m$  then  $\mathbb{TC}_{c_i} \leftarrow \mathbb{TC}_{c_i} \cup TC$ ;
16 return  $\mathbb{TC}_{c_i}$ ;
```

Algorithm 2: CMC Algorithm

```

1  $TC \leftarrow$  Sort clusters  $TC_i = (O_i, [l_i, r_i])$  from all cameras by  $l_i$ ;
2 for each  $TC_i$  associated with camera  $c_u$  do
3   for each candidate pattern  $\langle O_b, P_b \rangle \in Buf_{c_u}$  do
4     if  $|O_b \cap O_i| \geq m$  then
5       if  $|P_b| \geq k - 1$  then
6          $\mathbb{R} \leftarrow \mathbb{R} \cup \langle O_b \cap O_i, P_b \rightarrow c_u \rangle$ ;
7         for each camera  $c_o \in V_{out}^u$  do
8            $Buf_{c_o} \leftarrow Buf_{c_o} \cup \langle O_b \cap O_i, P_b \rightarrow c_u \rangle$ ;
9         for each camera  $c_o \in V_{out}^u$  do
10           $Buf_{c_o} \leftarrow Buf_{c_o} \cup \langle O_i, c_u \rangle$ ;
11 Remove dominated patterns in  $\mathbb{R}$ ;
12 return  $\mathbb{R}$ ;
```

a pattern from t timestamps to $t + 1$ timestamps. For each candidate pattern $\langle O_b, P_b \rangle$ in the local buffer of camera c_u , we examine the size of $O_b \cap O_i$. If $|O_b \cap O_i| \geq m$, we generate a new pattern $\langle O_b \cap O_i, P_b \rightarrow c_u \rangle$ and propagate it to the outgoing neighbor cameras in V_{out}^u . If this is a valid pattern, we also push it into result set \mathbb{R} . The algorithm terminates once all temporal clusters have been evaluated. Finally, we remove the dominated patterns in \mathbb{R} by building an inverted index to quickly examine the subset relationship between object sets.

4.2 Apriori Based Enumerator

Our second baseline is inspired by SPARE [6], which is a parallel and general framework to mine various co-movement patterns. Within SPARE, an Apriori-based enumerator is introduced to facilitate space pruning. To adopt the idea, we construct the combinatorial space using cameras because we can leverage the ontology of camera network to avoid enumerating all the combinations of cameras. In addition, it is straightforward that the anti-monotonicity property also holds:

PROPERTY 1. Let \mathbb{O}_i be the set of ϵ -reachable objects in all the cameras of P_i . If P_1 is a sub-path of P_2 , then $\mathbb{O}_1 \supseteq \mathbb{O}_2$.

With the property, if we cannot find a group of objects with size m that are ϵ -reachable in all cameras of P_i , then all the expanded paths from P_i can be pruned.

We present our Apriori-based enumerator in Algorithm 3. The general idea is to enumerate candidate paths in increasing number of cameras, evaluate their validity by assessing the constraints related to object size m and spatial proximity ϵ , and eliminate any illegal or dominated outcomes. The algorithm starts from the first layer to construct candidate patterns with only one camera. In the initialization stage (lines 1-4), temporal clustering via Algorithm 1 is employed to perform temporal clustering within each camera c_i . As long as its output \mathbb{TC}_{c_i} is not empty, the camera is considered as a candidate path with $len = 1$ and pushed into a buffer Q . These candidate cameras constitute the bottom layer of the lattice for Apriori-based enumeration.

In the subsequent layers, we adhere to the Apriori algorithm's procedure to enumerate candidate paths of length $n + 1$ derived from their two sub-paths of length n . Given $P = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n \rightarrow c_u$, we impose the condition that its two sub-paths, $P_1 = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n$ and $P_2 = c_2 \rightarrow \dots \rightarrow c_n \rightarrow c_u$, must be present in the output of the previous layer (lines 6-9). If either of these sub-paths is absent, path P can be pruned in accordance with the anti-monotonicity property. We construct \mathbb{O}_{new} to maintain a group of valid object sets that appear in both P_1 and P_2 , i.e., we intersect each pair of object set $O_1 \cap O_2$ and discard it if its size is smaller than m or it is dominated by other object set (lines 11-13). If \mathbb{O}_{new} is not empty, we find a new candidate path $P_{new} = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n \rightarrow c_u$ and push $(P_{new}, \mathbb{O}_{new})$ into Q . If $|P| \geq k$, we also add the valid candidate patterns into result set \mathbb{R} . When the buffer Q becomes empty, the Apriori-based enumerator terminates and we remove dominated patterns in \mathbb{R} . This step is the same as the one employed in the CMC algorithm.

Algorithm 3: Apriori-Based Enumerator

```

1 for each camera  $c_i$  do
2   Perform temporal clustering in  $c_i$  using Algorithm 1;
3    $P_i \leftarrow c_i; \mathbb{O}_i \leftarrow \mathbb{TC}_{c_i}$ ;
4   Add  $(P_i, \mathbb{O}_i)$  into buffer  $Q$ ;
5 while  $Q$  is not empty do
6   for each path  $P_1 = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n \in Q$  do
7     for each  $c_o \in V_{out}^n$  do
8       Let  $P_2 = c_2 \rightarrow c_3 \rightarrow \dots \rightarrow c_n \rightarrow c_o$ ;
9       if  $Q$  contains  $P_2$  then
10         $\mathbb{O}_{new} \leftarrow \{O_1 \cap O_2 | O_1 \in \mathbb{O}_1, O_2 \in \mathbb{O}_2\}$ ;
11        if  $|O_i| < m$  or  $O_i \subseteq O_j \in \mathbb{O}_{new}$  then
12          Remove  $O_i$  from  $\mathbb{O}_{new}$ ;
13        if  $|\mathbb{O}_{new}| > 0$  then
14           $P_{new} = P_1 \rightarrow c_o$ ;
15           $Q' \leftarrow Q' \cup (P_{new}, \mathbb{O}_{new})$ ;
16          if  $|P_{new}| \geq k$  then
17            for each  $O_i \in \mathbb{O}_{new}$  do
18               $\mathbb{R} \leftarrow \mathbb{R} \cup \langle P_{new}, O_i \rangle$ ;
19         $Q \leftarrow Q'; Q' \leftarrow \emptyset$ ;
20 Remove dominated patterns in  $\mathbb{R}$ ;
21 return  $\mathbb{R}$ ;
```

5 SEQUENCE-AHEAD MINING FRAMEWORK

In this section, we first deliver the basic idea of sequence-ahead mining framework and the construction of temporal-cluster suffix tree (TCS-tree) to produce co-movement pattern candidates. Among the candidates, we also propose an efficient sliding-window based verification algorithm to remove false positives and a hashing-based dominance elimination strategy to retain maximal patterns. At the end of the section, we discuss the extended support of the current data model for the scenario of overlapped cameras.

5.1 Basic Idea

A co-movement pattern involves three constraints that allow us to devise pruning rules, including the ϵ -reachable constraint for spatial proximity, the minimum group size m and the minimum path length k . The two baseline algorithms adopt different strategies to construct the search space for candidate enumeration. More specifically, the Apriori-based enumerator can be viewed as camera-ahead exploration and its search space consists of all the possible paths with length k . During the candidate enumeration, the constraints on parameters ϵ and m are further utilized for pruning. The CMC algorithm is temporal-ahead because it explores the search space from the temporal dimension. For each appearance of an object within a camera, it joins with the candidate patterns that are ϵ -reachable. The validity is further verified using constraints on parameters m and k .

In this paper, we propose a sequence-ahead mining framework with the objective of utilizing multiple constraints for candidate enumeration. A straightforward solution is to model each object o_i as a sequence of camera ids in ascending order of the entrance time into the camera. Then, we can utilize the constraints of minimum group size m and minimum path length k simultaneously, and perform frequent subsequence mining (FSM) for candidate generation. A subsequence is a candidate path only if it satisfies two conditions, namely its length is at least k , and its support is at least m . Since FSM is a fundamental data mining problem that has been adequately addressed, we can readily adopt existing frequent subsequence miners for candidate enumeration. For example, we can construct a suffix-tree for the sequences and use the index to identify frequent subsequences. For each candidate path, we can further apply the constraint of ϵ -reachability to eliminate false positives. Finally, the patterns that are not dominated will be returned to the query user as mining results.

The aforementioned solution still has great room for improvement. In the following, we first present the concept of meta-clusters and propose a new index called temporal-clustering suffix tree (TCS-tree) intends to integrate the three constraints for filtering in the frequent subsequence mining process. To reduce redundant verification overhead, we also propose a sliding-window based candidate enumeration strategy and a hashing based dominance eliminator.

5.2 TCS-tree Construction

In order to further refine the efficacy of the sequence-ahead mining framework and minimize the occurrence of false positive candidates, we propose an innovative strategy that involves extending the sequence of camera ids to incorporate temporal information and generate a more nuanced representation. The original sequence

is constructed using the one-to-one mapping function $F(o_i) = c_j$, with each element in the sequence representing a camera id. However, this representation is limited in that it is unable to differentiate between two distinct travel paths that share the same sequence of camera ids but possess substantially different starting times. To address this limitation, we intend to partition the temporal dimension in a camera c_j into multiple disjoint intervals I_j^1, I_j^2, \dots , and extend $F(o_i)$ to a new one-to-one mapping $G(o_i) = I_j^t$, thereby representing each object as a sequence of I_j^t . This approach enables us to incorporate the three constraints of ϵ -reachability, minimum group size m , and minimum path length k during the process of frequent subsequence mining. Notably, it is possible for an object o_i to be captured at multiple time periods at the same camera c_j . In this case, we can simply use o_i^1, o_i^2, \dots to represent different occurrence periods of o_i such that each o_i^u becomes the input of mapping function $G(\cdot)$. For ease of presentation, we still use $G(o_i)$ as the default notation unless further clarification is required.

It is obvious that smaller intervals are more powerful to distinguish false positives, i.e., if two objects share a travel path $P_i = c_1 \rightarrow \dots \rightarrow c_n$ but they are not ϵ -reachable, it is more likely for these two objects to be projected into different intervals. Therefore, we formulate the construction of function $G(o_i)$ as minimizing the total size of all intervals $\sum_{j,t} |I_j^t|$, with the following two constraints:

- (1) $G(o_i)$ is a one-to-one mapping in camera c_j , i.e., each object o_i is uniquely mapped to an interval I_j^t at camera c_j .
- (2) If o_u and o_v are ϵ -reachable in camera c_j , they are mapped to the same I_j^t . This requirement guarantees that there is no missing valid pattern.

Note that we cannot directly apply the temporal clustering algorithm in Algorithm 1 because it is possible that an object belongs to multiple clusters. This violates the constraint of one-to-one mapping requirement for $G(o_i)$ and we cannot represent an object with a sequence of cluster ids and then apply frequent subsequence mining. To resolve the issue, we propose a two-level temporal clustering algorithm and represent each object as a sequence of cluster ids, based on which we build a temporal cluster suffix tree (TCS-tree) for frequent subsequence mining.

In the first level of clustering, we apply Algorithm 1 to generate temporal clusters that are ϵ -reachable. In the second level, we progressively merge the temporal clusters that are overlapped. Two temporal clusters TC_i and TC_j are merged into a meta-cluster as long as they contain common objects, i.e., $TC_i.O_i \cap TC_j.O_j \neq \emptyset$. In implementation, we can scan the temporal clusters along the temporal dimension and treat the first cluster as a meta-cluster. When we access a temporal cluster TC_i , it is merged into previous meta-cluster MC if they are overlapped. Otherwise, TC_i and MC are disjoint and we treat TC_i as a new meta-cluster to merge its subsequent temporal clusters. The process terminates when all clusters have been examined.

Example 5.1. We use a toy example to explain the results of meta-clusters and the construction of TCS-tree. As depicted in Figure 2, there are 5 objects that move in a road network with cameras c_1 to c_6 . Each object is captured by part of these cameras. To construct TCS-tree, we can perform temporal clustering within each camera and merge the overlapped clusters into a meta-cluster. For example,

in camera c_5 , there are 3 temporal clusters $\{o_1, o_2, o_3\}, \{o_3, o_4\}, \{o_5\}$ and we can merge them into two meta-clusters $\{o_1, o_2, o_3, o_4\}, \{o_5\}$. Each object is uniquely mapped to a meta-cluster MC_j^i in camera c_i . Therefore, as shown in the figure, we can construct a new data model for each object and represent it as a sequence of meta-clusters. Then, we can build a TCS-tree on top of these sequences whose elements are meta-clusters. TCS-tree is essentially a suffix tree. Each unique suffix in the meta-cluster sequences is stored as a single node in a suffix tree. Each leaf node contains the starting position of the suffix it represents. Originally, there are 8 frequent subsequences. With the fine-grained representation, the number is reduced to 4.

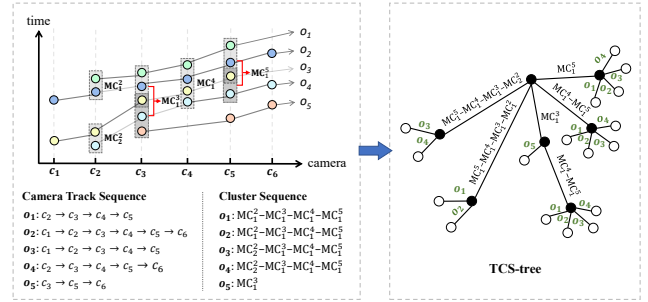


Figure 2: A toy example of the construction of TCS-Tree.

In the following, we analyze the properties of $G(o_i)$, which maps an object o_i into a meta-cluster generated by c_j . First, the mapping is one-to-one because the temporal dimension is split into disjoint meta-clusters and these meta-clusters cover all objects captured by the camera. In this way, each object can only be assigned to one of the meta-clusters. Second, for any two objects o_1 and o_2 that are ϵ -reachable, they will appear in the same meta-cluster. This is also straightforward because the temporal gap between two disjoint meta-clusters is at least ϵ . If o_1 and o_2 are located at different meta-clusters, they are unable to be ϵ -reachable.

Finally, we can also show that the meta-clusters generate the minimum interval size.

LEMMA 5.2. *The meta-clusters minimize $\sum_{j,t} |I_j^t|$.*

PROOF. We prove this by contradiction. Suppose there exists another mapping function $G'(o_i)$ that generates smaller intervals than the meta-clusters. Then, we can find a pair of objects (o_i, o_j) that belong to the same meta-cluster MC , but $G'(o_i) \neq G'(o_j)$. If o_i and o_j are ϵ -reachable, we finish the proof because the new mapping contradicts with the requirement that $G'(o_i) = G'(o_j)$ if o_i and o_j are ϵ -reachable. If o_i and o_j are not directly ϵ -reachable, based on the construction of meta-clusters, we can find a sequence of objects $S = o_i \rightarrow o_{i1} \rightarrow o_{i2} \rightarrow \dots \rightarrow o_j$ such that the neighboring objects are ϵ -reachable. Suppose $o_u \rightarrow o_v$ is the first pair of objects in S satisfying $G'(o_i) = G'(o_u) \neq G'(o_v)$. Then, we find a pair of objects (o_u, o_v) that are ϵ -reachable but $G'(o_u) \neq G'(o_v)$. This also leads to contradiction for mapping function $G'(\cdot)$. \square

5.3 Optimized Verification and Dominance Elimination

Assisted by TCS-tree, we can perform frequent subsequence mining to quickly identify candidate paths with at least k cameras and traversed by at least m objects. To support efficient verification on these candidates, we propose a sliding-window based candidate enumeration strategy and a hashing-based dominance eliminator.

5.3.1 Sliding-window Based Candidate Enumeration. Let $MC_1 \rightarrow MC_2 \rightarrow \dots \rightarrow MC_n$ ($n \geq k$) be a candidate sequence of meta-clusters generated by TCS-tree. Each meta-cluster MC_i is derived from a camera c_i and merged by multiple temporal clusters. An unoptimized verification procedure has to check all combinations of temporal clusters from different cameras, perform object intersection among the temporal clusters, and preserve result sets with at least m objects. The idea of CMC algorithm can be adopted to improve efficiency. We can traverse the cameras and maintain a buffer to store partial intersection results. When we access camera c_i , we only need to perform set intersection between the temporal clusters in c_i and the candidate object sets in the buffer. Nonetheless, the procedure requires heavy object set intersection operators. Even though the operation can be optimized to $O(m \log(n/m))$ ($m \leq n$) [1], this requires non-trivial implementation skills. In addition, the temporal clusters are overlapped with high redundancy. If we can save cost from unnecessary computation cost incurred by the redundant data, the performance can be further improved.

Algorithm 4 depicts our sliding-window based verification algorithm. The sequence of meta-clusters $MC_1 \rightarrow MC_2 \rightarrow \dots$ are processed in order. Let MC_i be the current meta-cluster in camera c_i . It essentially consists of a sequence of objects ordered by the entrance time to the camera and these objects are grouped into temporal clusters. We access these objects in order and maintain a sliding window $W = [o_l, o_r]$ using the data structure of queue, where o_r is the current object and o_l is the first object that belongs to the same temporal cluster with o_r . Then, enqueue and dequeue operations are used for adding or removing an element from the sliding window, respectively. We also maintain \mathbb{C} to store the candidate patterns generated during the verification process. For each candidate pattern $CP_i = \langle O_i, P_i \rangle$, we construct a queue $Q[CP_i]$ to store the common objects in O_i and the objects in the sliding window $[o_l, o_r]$. The elements in $Q[CP_i]$ will be dynamically updated when the sliding window moves.

When we access the current object o_r , we retrieve the set of candidate patterns in \mathbb{C} whose object set contains o_r . This step can be efficiently conducted by building an inverted index for the candidate patterns in an online fashion. For each pattern CP_i containing o_r , we append o_r to $Q[CP_i]$ with complexity $O(1)$. If o_r is the end of a temporal cluster and $|Q[CP_i]| \geq m$, we find and store the new candidate pattern $\langle Q[CP_i], P_i \rightarrow c_i \rangle$. The sliding window is also updated when o_r switches to a new temporal cluster. We iteratively remove o_l from the sliding window until o_l and o_r belong to the same temporal cluster. When o_l is removed from the sliding window, we also dequeue o_l from $Q[CP_i]$ for those candidate patterns CP_i containing o_l .

Example 5.3. We present an example in Figure 3 to explain the procedure of sliding-window based verification. In this example, we set $m = 2, k = 3$ and group objects within the same temporal cluster. Suppose we are now processing meta-cluster MC_4 and the candidate patterns after processing MC_1, MC_2 and MC_3 include CP_1 and CP_2 . The objects in MC_4 are accessed in the order of $o_3 \rightarrow o_4 \rightarrow o_2 \rightarrow o_5$ and they form two temporal clusters $\{o_3, o_4, o_2\}$ and $\{o_4, o_2, o_5\}$. When the first two objects o_3 and o_4 are accessed, the sliding window is expanded to $[o_3, o_4]$. By checking the inverted index, we also append o_3 to $Q[CP_1]$ and o_4 to $Q[CP_2]$. When o_2 is accessed, it is enqueued into the sliding window, which becomes $[o_3, o_4, o_2]$. Since o_2 is contained in CP_2 , we add o_2 to $Q[CP_2]$, resulting in a new candidate pattern $CP' = \langle \{o_4, o_2\}, c_2 \rightarrow c_3 \rightarrow c_4 \rangle$. Finally, when the last object o_5 is accessed, o_3 is dequeued from the sliding window and o_5 is enqueued. $Q[CP_2]$ is updated and another new candidate pattern $CP'' = \langle \{o_4, o_2, o_5\}, c_2 \rightarrow c_3 \rightarrow c_4 \rangle$ is generated.

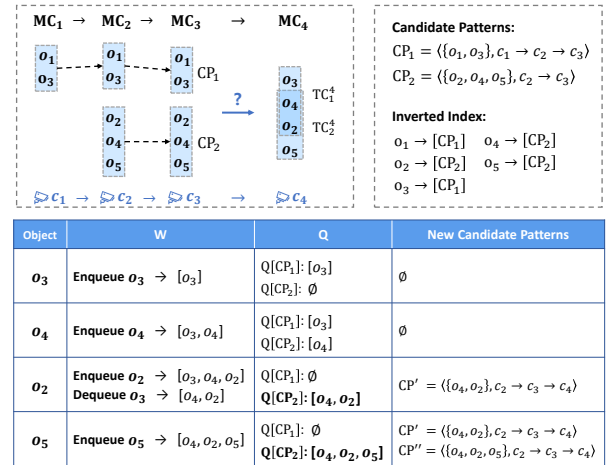


Figure 3: An example of Sliding-window Based Verification

Compared with traditional CMC algorithm that requires frequent set intersection, our sliding-window based verification strategy has two advantages. First, we require no explicit set intersection. The objects are enqueued and dequeued with $O(1)$ complexity to update the common objects shared by the sliding window and candidate patterns. Second, each object is accessed only once, even when it is contained in multiple temporal clusters. This property is useful to handle overlapped temporal clusters with high redundancy and can save considerable computation cost.

5.3.2 Hashing-based Dominance Verification. In order to identify all the maximal patterns, we need to perform dominance verification among the co-movement patterns in \mathbb{R} and eliminate those that are dominated. A naive approach is to examine each $CP_i = \langle O_i, P_i \rangle \in \mathbb{R}$ whether there exists a pattern $CP_j = \langle O_j, P_j \rangle \in \mathbb{R}$ that dominates it, by verifying whether $O_i \subseteq O_j$ and $P_i \subseteq P_j$. However, this approach is inefficient and entails a considerable amount of unnecessary examination cost when CP_i is not dominated by CP_j .

In order to enhance efficiency in dominance verification, we put forward a hashing-based approach. Rather than comparing each

Algorithm 4: Sliding-window Based Verification

```
1  $\mathbb{C} \leftarrow \emptyset$ ;  
2 for each  $MC_i$  with camera  $c_i$  in the frequent sequence do  
3    $W \leftarrow \text{queue}$ ;  $Q \leftarrow \emptyset$ ;  $\mathbb{C}_{new} \leftarrow \emptyset$ ;  
4   for each candidate pattern  $CP_i \in \mathbb{C}$  do  
5      $Q[CP_i] \leftarrow \text{queue}$ ;  
6    $S \leftarrow$  objects of  $MC_i$  in order of entrance time in  $c_i$ ;  
7   for each object  $o_r$  in  $S$  do  
8     Enqueue  $o_r$  into  $W$ ;  
9     for each  $CP_i \in \mathbb{C}$  containing  $o_r$  do  
10      Enqueue  $o_r$  into  $Q[CP_i]$ ;  
11      if  $o_r$  is the end anchor of a temporal cluster then  
12        for each  $CP_i \in \mathbb{C}$  do  
13          if  $|Q[CP_i]| \geq m$  then  
14             $\mathbb{C}_{new} \leftarrow \mathbb{C}_{new} \cup \langle Q[CP_i], P_i \rightarrow c_i \rangle$ ;  
15            while  $o_l$  and  $o_r$  in different temporal clusters do  
16              Dequeue  $o_l$  from  $W$  and  $Q$ ;  
17          for each candidate pattern  $CP_i \in \mathbb{C}$  do  
18            if  $|P_i| \geq k$  then  $\mathbb{R} \leftarrow \mathbb{R} \cup CP_i$ ;  
19          for each  $TC_j \in MC_i$  in the frequent sequence do  
20             $\mathbb{C}_{new} \leftarrow \mathbb{C}_{new} \cup \langle TC_j, c_i \rangle$ ;  
21    $\mathbb{C} \leftarrow \mathbb{C}_{new}$ ;  
22 return  $\mathbb{R}$ ;
```

CP_i against all other patterns in \mathbb{R} , we need only examine whether CP_i is dominated by CP_j with $O_i = O_j$ or $P_i = P_j$, by virtue of the following lemma:

LEMMA 5.4. *For any pattern $CP_i = \langle O_i, P_i \rangle$ in the result set \mathbb{R} generated by Algorithm 4, if CP_i is dominated in \mathbb{R} , then there must exist another pattern $CP = \langle O, P \rangle \in \mathbb{R}$ such that $O_i = O$ or $P_i = P$.*

PROOF. We prove by contradiction. Let us assume CP_i is dominated by a pattern $CP_j = \langle O_j, P_j \rangle \in \mathbb{R}$ and we have $O_i \subseteq O_j$ and $P_i \subseteq P_j$. If there is no pattern $CP = \langle O, P \rangle \in \mathbb{R}$ with $O_i = O$ or $P_i = P$, we can conclude that $O_i \neq O_j$ and $P_i \neq P_j$ (i.e., $O_i \subset O_j$ and $P_i \subset P_j$).

Without loss of generality, we denote $P_i = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n$. Since $CP_j = \langle O_j, P_j \rangle$ is a valid pattern, there exists a temporal cluster with object set $O_c \supseteq O_j$ for camera c_n . On the other hand, since O_i is a pattern in \mathbb{R} , we assume it is generated by the candidate pattern $CP_u = \langle O_u, P_u \rangle$ with $O_u \supseteq O_i$ and $P_u = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_{n-1}$. Since O_c is also a temporal cluster in camera c_n , we know from Algorithm 4 that O_c and candidate pattern CP_u can lead to a new pattern $\langle O_c \cap O_u, P_i \rangle$, which dominates CP_i because $O_i \subseteq O_c \cap O_u$ and their travel path is identical. The pattern $\langle O_c \cap O_u, P_i \rangle$ will be added into \mathbb{R} and thus, we find a pattern leading to a contradiction. \square

In this way, we can build two hash maps HTP and HTO . HTP uses the travel path as the key and HTO uses the object set as the key. For each candidate pattern CP_i , we only need to compare with the patterns in $HTP[P_i]$ and $HTO[O_i]$.

5.4 Complexity Analysis

Frequent subsequence mining. To build TCS-tree, temporal clustering is performed within each camera with linear complexity.

Given N objects with average travel path L , the complexity is $O(NL)$. Since the merge operation to generate meta-clusters, the construction of suffix tree, and applying suffix tree for frequent subsequence mining are all linear [23], the total complexity in this module is $O(NL)$.

Sliding-window based verification. Given M frequent subsequences generated from TCS-tree, we need to perform the sliding-window based verification for each sequence of meta-clusters, whose average length is assumed to be \bar{k} . From Algorithm 4, we know that the processing cost of each meta-cluster (lines 3-21) is linearly correlated with inverted index size and meta-cluster size. Suppose there are b candidate patterns in the buffer and each contains an average of \bar{m} objects. Let c be the average size of a meta-cluster. Then, the cost of sliding-window based verification is $O(M\bar{k}(b\bar{m} + c))$.

Hashing-based dominance eliminator. The complexity of this component is straightforward to estimate. For each candidate CP_i , we only need to compare with the patterns in $HTP[P_i]$ and $HTO[O_i]$. So the complexity is determined by the length of $HTP[P_i]$ and $HTO[O_i]$ and the examination cost of subset relationship.

5.5 Extension to Overlapped Cameras

In real-world surveillance camera systems, a moving object could be captured by multiple cameras simultaneously, especially in the area of road intersections. To handle the scenario, our idea is to cluster the cameras with overlapped view into a virtual camera V . For any camera $c_i \in V$, there exists $c_j \in V$ such that c_i and c_j have overlapped view. For any two cameras from different clusters, they are not overlapped. Suppose a moving object is captured by multiple cameras within V , which we denote as $(c_i, [s_i, e_i]), \dots, (c_j, [s_j, e_j])$ and these time intervals are overlapped. We can replace them as $(V, [\min(s_i, \dots, s_j), \max(e_i, \dots, e_j)])$. Hence, the travel path of each object is represented as a sequence of camera cluster IDs and the enlarged time intervals. In this way, the representation of travel path is consistent with our original data model and we can still apply the proposed mining algorithms to identify co-movement patterns of grouped objects.

6 EXPERIMENTS

6.1 Experimental Setup

Datasets. Since there lack large-scale and publicly-accessible trajectories recovered from videos, we construct two types of datasets for co-movement pattern mining. In the first group of datasets, we use real GPS trajectories and road network to generate approximate trajectories recovered from videos. We can deploy a specified number of cameras randomly on the road network. The position and view field of each camera are considered as prior knowledge. From the GPS trajectories, we can roughly estimate the entrance time and exit time from a camera according to the travel speed estimation and the attributes of the camera. In this way, we can convert a GPS trajectory into a sequence of camera IDs and the time interval captured by each camera is also available. In our implementation, we use real trajectory datasets of DIDI Chengdu [22] and Singapore Taxi [6] as well as their road network to construct two video trajectories for co-movement pattern mining.

In the second group of dataset, we start from raw videos and employ the end-to-end mining pipeline presented in Section 2.3

to evaluate the effectiveness of co-movement pattern mining from videos. We adopt Carla [9] to generate synthetic video benchmark datasets. In this dataset, we generate 1169 cameras, with 1833 vehicles moving across the road network. We apply recent trajectory recovery algorithm [27] on these raw videos to generate a sequence of camera id for each vehicle. As the counterpart, we can also derive the GPS trajectories of vehicles from the simulation engine of Visual Road. These GPS trajectories will be used as groundtruth for effectiveness evaluation.

Comparison Approaches. We compare the proposed **TCS-tree** with the two baselines derived from previous co-movement pattern miners, namely **CMC** and **Apriori**. In addition, we also incorporate the baseline presented in Section 5.1. The algorithm, denoted by **FSM**, adopts suffix tree and our proposed sequence-ahead mining framework to mine frequent subsequences. It does not leverage our improved strategies proposed in Section 5.

Parameter Setup. We examine the scalability of the proposed algorithms in terms of the six parameters listed in Table 3. Among them, k , m , ϵ are related to the definition of a co-movement pattern. The remaining parameters are related to dataset settings, including moving object number, camera number and the average travel length per object. We use the same settings for query parameters k , m and ϵ for pattern mining in Chengdu and Singapore datasets. However, the parameters related to dataset settings are different. Overall, the Singapore dataset has a much smaller number of distinct objects, but with longer travel path per vehicle. The default parameters are highlighted in bold.

Table 3: Parameter setup for scalability analysis.

Query path length k	Singapore	2, 3, 4, 5, 6, 7, 8
	Chengdu	
Group size m	Singapore	2, 3, 4, 5, 6, 7, 8
	Chengdu	
Threshold ϵ	Singapore	40, 50, 60 , 70, 80, 90, 100
	Chengdu	
Camera number	Singapore	10k, 15k, 20k, 25k, 30k, 35k
	Chengdu	3k, 4k, 5k, 6k, 7k, 8k, 9k
Object number	Singapore	1.5k, 2.1k, 2.3k, 2.5k, 2.7k
	Chengdu	6k, 7k, 8k, 9k, 10k, 11k, 12k
Average path length (#cameras per vehicle)	Singapore	200, 300, 400, 500, 600, 700, 800
	Chengdu	90, 100, 110, 120, 130, 140, 150

All the algorithms are implemented in C++ language and conducted on the server with a 3.20 GHz i9-12900K CPU, 128GB of main memory and 1 TB hard drive. The server runs a Ubuntu Linux release 20.04.5.

6.2 Scalability Analysis

Increasing path length k . In the first experiment, we report the latency of co-movement pattern mining w.r.t. increasing k . As shown in Figure 4, the two baselines are inefficient, with their execution time being at least 100 times greater than that of our proposed TCS-tree approach. They need to enumerate and verify an enormous number of candidate patterns, necessitating frequent operations of object set intersection. To investigate this phenomenon further, we examine the frequency of object set intersections incurred during

the pattern mining process. The findings reveal a direct correlation between this indicator and the overall execution time. Apriori and CMC trigger billions of set intersection operations, which become the primary performance bottleneck. Moreover, we observe that these two baselines are not sensitive to k . Since the mining task intends to find non-dominated maximal patterns, they need to enumerate all the candidate paths with varying length.

The FSM algorithm incorporates our proposed sequence-ahead mining framework, surpassing the performance of the baselines and affirming the effectiveness of leveraging suffix trees to filter out false positive candidates. Moreover, as k increases, both FSM and TCS-tree demonstrate enhanced pruning capabilities. The frequent subsequence mining technique expeditiously filters a substantial number of candidate patterns with lengths smaller than k . Our TCS-tree method achieves a significant improvement over FSM. Its index is based on sequences of meta-clusters, which are more fine-grained than the camera sequences in FSM. In addition, it incorporates a sliding-window-based verification and a hashing-based dominance eliminator to further improve efficiency.

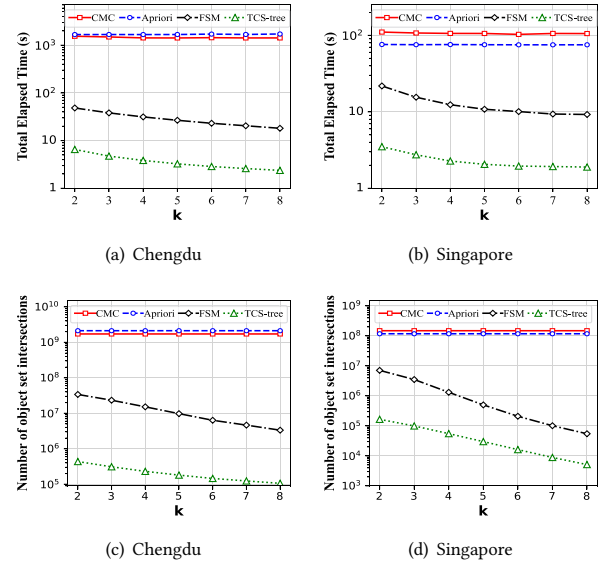


Figure 4: Increasing query path length k .

Increasing group size m . The results with increasing group size m are reported in Figure 5. It is interesting to find that CMC and Apriori are very sensitive to parameter m . When m is large, the number of candidate patterns undergoes a significant reduction, enabling them to even surpass the performance of the FSM algorithm. This behavior can be attributed to the fact that CMC and Apriori retain candidate patterns comprising a minimum of m objects. When m increases, the number of candidate patterns is sharply reduced, resulting in much lower computation cost for object set intersection. For FSM and TCS-tree, their performances initially benefit from the increase of m because fewer frequent subsequences are generated. When m continues to grow, there exist a small number of candidate patterns and the total computation overhead is dominated by the index construction and frequent subsequence mining.

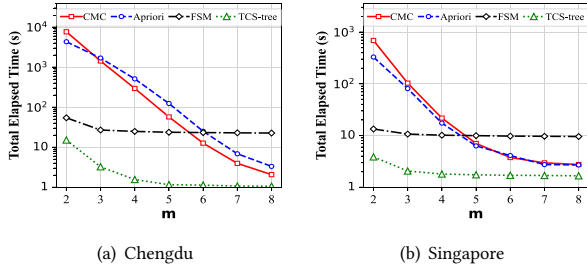


Figure 5: Increasing m .

Increasing proximity threshold ϵ . As ϵ increases, there are more valid candidate patterns, leading to an escalation in the running time of all the algorithms. Notably, the FSM algorithm is less sensitive to ϵ because it does not leverage the parameter for candidate pruning in the stage of frequent subsequence mining.

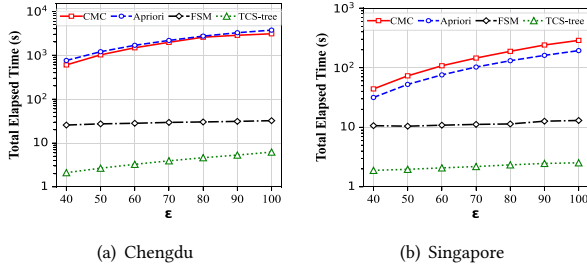


Figure 6: Increasing ϵ .

Increasing dataset-related parameters. In Figure 7, we perform scalability analysis w.r.t. concerning the number of cameras, moving objects, and the average length of travel paths for each object. It's not surprising to observe that the figures illustrate an evident ascending trend across all the algorithms.

Notably, the two baselines display a greater sensitivity to the cardinality of the dataset, including the number of objects and the travel path length of each object. They exhibit a considerably more pronounced increasing trend in the two factors when compared with the factor of camera number. Even though CMC applies Apriori algorithm to the combination of cameras, the road network ontology remains unchanged. Hence, the deployment of additional cameras in the road network does not yield a significant increase in the number of edges within the derived camera network.

6.3 Break-down Analysis and Ablation Study

This experiment is designed with two goals. Firstly, considering that the TCS-tree algorithm comprises three essential steps, including frequent subsequence mining, verification of candidate patterns, and the elimination of dominated patterns, we conduct a break-down analysis to assess the computational cost associated with each component. Secondly, our technical contributions originate from TCS-tree construction, sliding-window-based verification, and hashing-based dominance elimination, so we perform an ablation

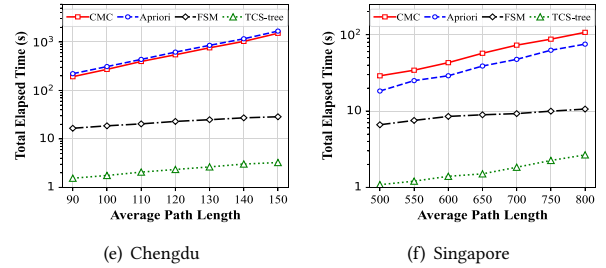
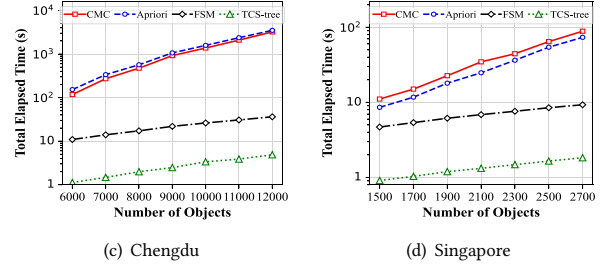
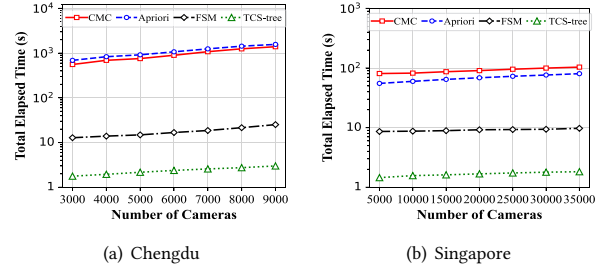


Figure 7: Adjusting dataset-related parameters.

study to evaluate their impact by devising three variants of our proposed algorithm. The first variant, denoted as TCS-v1, replaces the hashing-based dominance eliminator with the one employed in the baseline methods. The second variant, known as TCS-v2, substitutes the sliding-window-based verification with CMC algorithm. The third variant, namely TCS-v3, builds TCS-tree on top of sequences of camera ids instead of meta-clusters and applies our proposed verification and de-dominance techniques. The goal is to identify the pruning power of fine-grained representation for frequent subsequence mining.

We run these algorithms with default parameter settings in Table 3. From the running time results in Figure 8, the verification module incurs the highest computational cost, accounting for approximately 75% of the total processing time in the Chengdu dataset. Disabling the hashing-based dominance eliminator in TCS-v1 leads to an almost twofold cost increase in this component. Similarly, removing the sliding-window based verification in TCS-v2 results in a significant cost escalation. The fine-grained data model that represents each object as a sequence of meta-clusters plays the most important role. Substituting this model with sequences of camera IDs dramatically increases the expense of frequent subsequence mining. The reason is that in the original TCS-tree, meta-clusters with size smaller than m can be removed, yielding shorter sequences

compared to TCS-v3 and reducing the computational cost for suffix tree construction and frequent subsequence mining. TCS-v3 incurs higher costs for verification and dominance elimination due to its generation of a greater number of candidate patterns. To provide a baseline comparison, we also included FSM in the figure, which lacks the fine-grained data model, sliding-window based verification, and hashing-based dominance eliminator.

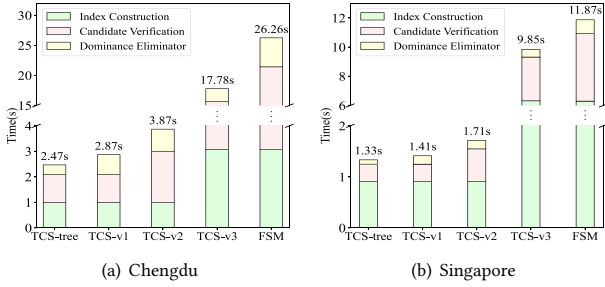


Figure 8: Break-down analysis and ablation study.

6.4 Effectiveness Analysis

In the final experiment, our goal is to examine the congruence between the co-movement patterns derived from video recordings and those generated by GPS trajectories. In our datasets generated by Carla, the GPS trajectories can be predetermined by the simulation engine, thereby enabling us to employ them as a reference to generate accurate patterns. Regarding the video data, we utilize the TCS-tree methodology on two distinct data sources. In the first data source, we adopt an existing map matching algorithm to convert the GPS trajectories into camera sequences, which are referred to as oracle travel paths. This is equal to assuming the existence of a perfect trajectory recovery algorithm that generates correct sequences for each moving object. In the second source, we implement the pipeline outlined in Section 2.3, wherein the TCS-tree approach is employed to identify co-movement patterns based on imperfect trajectories extracted from the videos.

We employ the flock pattern mining algorithm [24] to establish the groundtruth for co-movement patterns derived from GPS trajectories. The flock pattern algorithm necessitates the consideration of spatial proximity between pairs of objects, which aligns with our requirement for ϵ -reachability between object pairs. We utilize the F_1 -score as the performance metric. A co-movement pattern CP_1 generated by our algorithm is deemed to match CP_2 from the groundtruth patterns if they both encompass the same set of objects, while ensuring that the intersection over union (IoU) between their respective time spans exceeds the threshold of 0.8.

Regarding the parameters associated with flock patterns, we set the minimum group size as 2, the minimum time span as 10 seconds, and the disk radius as 100 meters. It takes 36.6 seconds to mine flock patterns from GPS data, whereas TCS-tree only consumes 0.2 seconds for co-movement pattern mining from the extracted camera sequences. As to pre-processing step for trajectory recovery from videos, it takes the method proposed in [27] 3.5 hours to perform iterative clustering on the snapshots captured by the

cameras. The objects within the same cluster are considered to be the same entity. Fortunately, this pre-processing step is executed only once. Furthermore, in addition to facilitating co-movement pattern analysis, the output of this step can be effectively leveraged to support diverse downstream applications.

As shown in Figure 9, we report the F_1 -scores with varying travel path length among the detected patterns. Notably, when executing the TCS-tree algorithm on the oracle camera sequences derived from GPS trajectories, the attained F_1 -scores are close to 1. This outcome suggests a remarkable similarity between co-movement pattern mining driven by video analysis and conventional GPS-driven approaches. Hence, in scenarios where GPS data is inaccessible, a sole reliance on video data sources for co-movement pattern mining emerges as a viable alternative. In practical applications, it is worth noting that trajectory recovery algorithms such as [27] inherently generate imperfect travel paths. Consequently, when employing the pipeline introduced in Section 2.3 on raw video footage, the F_1 -scores exhibit a noticeable decrease, yet they remain within an acceptable range.

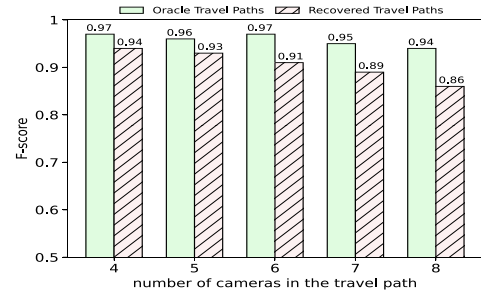


Figure 9: Effectiveness analysis for video-based co-movement pattern mining.

7 CONCLUSION AND FUTURE STUDY

The study at hand makes the first attempt at co-movement pattern mining from video data. We formulate the problem and theoretically prove it is NP-hardness. By adopting ideas from existing works, we develop two baselines. We also propose a novel index TCS-tree and an efficient verification strategy based on sliding window and a hashing-based dominance eliminator. Experimental results validate the efficiency of the proposed index and mining algorithm.

While this work is commendable, there remain certain limitations that indicate a wide range of untapped opportunities. Firstly, the focus of the present study is a specific case of co-movement patterns. Future research can explore diversified co-movement pattern definitions. Secondly, since the trajectories recovered from videos are not perfect, probabilistic pattern mining could be an interesting direction. Finally, we can explore co-movement pattern mining from alternative sensory data such as bluetooth data.

ACKNOWLEDGMENTS

The work is supported by the National Key Research and Development Project of China (2022YFF0902000) and CCF-Huawei Populus Grove Fund.

REFERENCES

- [1] Ricardo Baeza-Yates and Alejandro Salinger. 2010. Fast Intersection Algorithms for Sorted Sequences. In *Algorithms and Applications*. Springer, 45–61.
- [2] Alka Bhushan, Umesh Bellur, Kuldeep Sharma, Srijay Deshpande, and Nandlal L. Sarda. 2017. Mining swarm patterns in sliding windows over moving object data streams. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.
- [3] Lu Chen, Yunjun Gao, Ziquan Fang, Xiaoye Miao, Christian S. Jensen, and Chenjuan Guo. 2019. Real-time Distributed Co-Movement Pattern Detection on Streaming Trajectories. *Proc. VLDB Endow.* 12, 10 (2019), 1208–1220.
- [4] Yueting Chen, Nick Koudas, Xiaohui Yu, and Ziqiang Yu. 2022. Spatial and Temporal Constrained Ranked Retrieval over Videos. *Proc. VLDB Endow.* 15, 11 (2022), 3226–3239.
- [5] Yueting Chen, Xiaohui Yu, Nick Koudas, and Ziqiang Yu. 2021. Evaluating Temporal Queries Over Video Feeds. In *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*, Guoliang Li, Zhanhui Li, Stratos Dredo, and Divesh Srivastava (Eds.). ACM, 287–299.
- [6] Qi Fan, Dongxiang Zhang, Huayu Wu, and Kian-Lee Tan. 2016. A General and Parallel Platform for Mining Co-Movement Patterns over Large-scale Trajectories. *Proc. VLDB Endow.* 10, 4 (2016), 313–324.
- [7] Ziquan Fang, Yunjun Gao, Lu Pan, Lu Chen, Xiaoye Miao, and Christian S. Jensen. 2020. CoMing: A Real-time Co-Movement Mining System for Streaming Trajectories. In *SIGMOD*. ACM, 2777–2780.
- [8] Joachim Gudmundsson and Marc J. van Kreveld. 2006. Computing longest duration flocks in trajectory data. In *GIS*, Rolf A. de By and Silvia Nittel (Eds.). ACM, 35–42.
- [9] Brandon Haynes, Amrita Mazumdar, Magdalena Balazinska, Luis Ceze, and Alvin Cheung. 2019. Visual Road: A Video Data Management Benchmark. In *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*, Peter A. Boncz, Stefan Manegold, Anastasia Ailamaki, Amol Deshpande, and Tim Kraska (Eds.). ACM, 972–987.
- [10] Yuhang He, Jie Han, Wentao Yu, Xiaopeng Hong, Xing Wei, and Yihong Gong. 2020. City-Scale Multi-Camera Vehicle Tracking by Semantic Attribute Parsing and Cross-Camera Tracklet Matching. In *CVPR Workshops*. IEEE, 2456–2465.
- [11] Hung-Min Hsu, Tsung-Wei Huang, Gaoang Wang, Jiarui Cai, Zhichao Lei, and Jenq-Neng Hwang. 2019. Multi-Camera Tracking of Vehicles based on Deep Features Re-ID and Trajectory-Based Camera Link Models. In *CVPR Workshops*. Computer Vision Foundation / IEEE, 416–424.
- [12] Hoyoung Jeung, Man Lung Yiu, Xiaofang Zhou, Christian S. Jensen, and Heng Tao Shen. 2008. Discovery of convoys in trajectory databases. *Proc. VLDB Endow.* 1, 1 (2008), 1068–1080.
- [13] Yuxuan Li, James Bailey, and Lars Kulik. 2015. Efficient mining of platoon patterns in trajectory databases. *Data Knowl. Eng.* 100 (2015), 167–187.
- [14] Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays. 2010. Swarm: Mining Relaxed Temporal Moving Object Clusters. *Proc. VLDB Endow.* 3, 1 (2010), 723–734.
- [15] Zongyu Lin, Guozhen Zhang, Zhiqun He, Jie Feng, Wei Wu, and Yong Li. 2021. Vehicle Trajectory Recovery on Road Network Based on Traffic Camera Video Data. In *GIS*. ACM, 389–398.
- [16] Yiyang Liu, Hua Dai, Bohan Li, Jiawei Li, Geng Yang, and Jun Wang. 2021. ECMA: An Efficient Convoy Mining Algorithm for Moving Objects. In *CIKM*, Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong (Eds.). ACM, 1089–1098.
- [17] Faisal Orakzai, Toon Calders, and Torben Bach Pedersen. 2019. k/2-hop: Fast Mining of Convoy Patterns With Effective Pruning. *Proc. VLDB Endow.* 12, 9 (2019), 948–960.
- [18] Faisal Orakzai, Thomas Devogele, and Toon Calders. 2015. Towards distributed convoy pattern mining. In *SIGSPATIAL*, Jie Bao, Christian Sengstock, Mohammed Eunus Ali, Yan Huang, Michael Gertz, Matthias Renz, and Jagan Sankaranarayanan (Eds.). ACM, 50:1–50:4.
- [19] Thi Thi Shein, Sutteera Puntheeranurak, and Makoto Imamura. 2020. Discovery of evolving companion from trajectory data streams. *Knowl. Inf. Syst.* 62, 9 (2020), 3509–3533.
- [20] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David C. Anastasiu, and Jenq-Neng Hwang. 2019. CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification. In *CVPR*. Computer Vision Foundation / IEEE, 8797–8806.
- [21] Panrong Tong, Mingqian Li, Mo Li, Jianqiang Huang, and Xiansheng Hua. 2021. Large-scale vehicle trajectory reconstruction with camera sensing network. In *MobiCom*. ACM, 188–200.
- [22] Yongxin Tong, Yuxiang Zeng, Zimu Zhou, Lei Chen, Jieping Ye, and Ke Xu. 2018. A Unified Approach to Route Planning for Shared Mobility. *Proc. VLDB Endow.* 11, 11 (2018), 1633–1646.
- [23] Esko Ukkonen. 1995. On-line construction of suffix trees. *Algorithmica* 14, 3 (1995), 249–260.
- [24] Marcos R. Vieira, Petko Bakalov, and Vassilis J. Tsotras. 2009. On-line discovery of flock patterns in spatio-temporal data. In *ACM GIS*, Divyakant Agrawal, Walid G. Aref, Chang-Tien Lu, Mohamed F. Mokbel, Peter Scheuermann, Cyrus Shahabi, and Ouri Wolfson (Eds.). ACM, 286–295.
- [25] Yida Wang, Ee-Peng Lim, and San-Yih Hwang. 2006. Efficient mining of group patterns from user movement data. *Data Knowl. Eng.* 57, 3 (2006), 240–282.
- [26] Munkh-Erdene Yadamjav, Zhifeng Bao, Baihua Zheng, Farhana M Choudhury, and Hanan Samet. 2020. Querying recurrent convoys over trajectory data. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11, 5 (2020), 1–24.
- [27] Fudan Yu, Wenxuan Ao, Huan Yan, Guozhen Zhang, Wei Wu, and Yong Li. 2022. Spatio-Temporal Vehicle Trajectory Recovery on Road Network Based on Traffic Camera Video Data. In *KDD*. ACM, 4413–4421.