Article

Not peer-reviewed version

How Much to Learn? An Information-Sufficiency Criterion for Detecting Motion Rules in Scenario Surveillance

Herlindo Hernandez-Ramirez , Jorge Luis Perez-Ramos , Daniel Canton-Enriquez , Ana-Marcela Herrera-Navarro , Hugo Jimenez-Hernandez *

Posted Date: 6 October 2025

doi: 10.20944/preprints202510.0359.v1

Keywords: smart surveillance; automatic learning; motion dynamic patterns; modelling

 doi:10.20944/preprints202510.0359.v1

Article

How Much to Learn? An Information-Sufficiency Criterion for Detecting Motion Rules in Scenario Surveillance

Herlindo Hernandez-Ramirez [1], Jorge Luis Perez-Ramos [2], Daniel Canton-Enriquez [2], Ana Marcela Herrera-Navarro [2] and Hugo Jimenez-Hernandez [2,*]

[1] CIDESI, Av. Pie de la Cuesta No. 702, Desarrollo San Pablo, 76125 Santiago de Querétaro, Querétaro, México
[2] Fac. Informática UAQ, Av. de las Ciencias S/N, 76230 Juriquilla, Querétaro, México
* Correspondence: hugo.jimenez@uaq.mx; Tel.: +52-442-473-2064

Abstract

An automated learning process is a powerful tool that extracts useful data structures by measuring information, expressing conditions, and regulating invariant tolerance. Surveillance video systems and intelligent video analysis have become increasingly valuable, as it is often challenging to discern and represent the primary motion patterns necessary for automatic analysis. Combining automated learning with video analysis enables the development of intelligent analytic systems that can operate effectively in uncertain scenarios, automatically identifying dominant motion dynamics. The ability to recognize these motion dynamics relies on the theoretical framework used to represent them and the learning process employed to identify patterns. In the literature, a state-system-based approach offers a way to describe temporal and spatial interrelationships, which is essential for understanding motion dynamics. A key aspect of this approach is determining when the number of positive learned states from a given information source becomes sufficient to detect dominant motion in a surveillance video system. This determination is crucial as it impacts the variability of movements that the monitored subjects can exhibit while adhering to the camera's view projection and the resources required for implementation. While previous research has produced practical approaches, most have been limited to specific scenarios, underscoring the need for further investigation. In this study, we outline several advantages of establishing a formal criterion to determine when a symbolic system in a surveillance scenario has developed a robust and practical model to explain the observed motion dynamics. Our proposal is sustained by the hypothesis that a correct model can reliably account for most motion dynamics over time in an automatic learning process. The validation is performed with different real scenarios, where the central dynamic becomes learned, labeling those dynamics that the learned model cannot explain. In a real application, the approach was used to model traffic vehicles on the avenues of Queretaro City.

Keywords: smart surveillance; automatic learning; motion dynamic patterns; modelling

1. Introduction

In rapidly growing cities, it is common to implement and integrate low-cost video surveillance systems that can operate effectively in various environmental conditions, both indoors and outdoors. This generates a massive volume of information, which is impossible to analyze comprehensively. From this perspective, it is necessary to integrate intelligent systems that assist in processing this information, with an emphasis on detecting and analyzing events of interest [1,2]. The detection of these events is particularly relevant in applications aimed at crime prevention, tracking persons of interest [3], detecting atypical behavior [4,5], optimizing resources [6], and crowd dynamics analysis [7].

 doi:10.20944/preprints202510.0359.v1

The literature review identifies the main variables that affect the effectiveness of existing techniques, classifying them as follows: i) the quality of the data collected, since its accuracy and completeness directly impact the effectiveness of the process [8–10]; ii) the frequency of sampling data, referring to the periodicity with which the data is renewed, which significantly influences the ability to detect activities of interest [11]; iii) coverage or scope, which determines the susceptibility of events to be detected as a result of the perspective of the cameras and possible occlusions present in the scene associated with the semantic of the dynamic observed [12–14]; and finally, iv) the sensitivity of the system, referring to the detection threshold that allows us to distinguish between an event of interest and a possible false positive [15–17].

In this context, intelligent surveillance systems employ adaptive machine learning techniques that enable the identification and extraction of movement patterns, thereby representing the dynamics of the scene. For this purpose, the algorithms prioritize two aspects: the way in which movement is defined in the scene and the way in which the structures of that movement are modeled.

In the first step, the criteria for encoding information are analyzed so that, once the data of interest has been expressed, it can be processed in a compact manner consistent with the semantics of the objects' dynamics present in the scene. Under this criterion, some studies use segmentation masks that delimit the regions occupied by the object in the scene [18]. Other approaches consider the region of interest through silhouettes as binary masks or heat maps, where the objects of interest correspond to the pixels with the highest value. In contrast, irrelevant information approaches values close to zero[19–21]. In addition, there are innovative approaches based on deep neural networks that extract relevant attributes such as the position, aspect ratio, and height of objects [22,23]. The second criterion analyzes the structures that allow movement structures to be grouped and modeled, which in turn can be: (i) by perimeter delimitation of the geometry of the shape, which considers information about the texture or color of objects to identify them and enclose them in a space that can be rectangular, square, or elliptical [24], some based on object contours [25,26], these works implement length and area restrictions on the region to be segmented by means of cost functions, in particular [27] applies methods based on convolutional neural networks (CNN); (ii) using hierarchical structures based on enumeration, some representative works are [28,29], where they use multiscale inference approaches under the approach of combining multiscale predictions.

For cases where it is necessary to model the dynamics of the scene, allowing the detection of the dominant dynamics for the description of activities over time, this is of interest in automatic monitoring activities, where the objective is usually to recognize and differentiate the observed activities [30–32]. To this end, different techniques have been developed that consider information about movement in the scene as evidence to infer models capable of describing the observed activities.

The efficiency of these models depends entirely on the learning processes implemented, as the proper definition of movement rules in a scenario directly impacts the efficiency of monitoring and security systems [2]. For this reason, the ability to identify relevant movement patterns in a scenario is a fundamental task for detecting or omitting events that may be considered of interest in monitoring applications.

Currently, there are tasks for which no general solution exists [33,34]. Efforts are focused on solving challenges such as missing information resulting from random or deterministic factors, which motivates strategies based on techniques like k-nearest neighbors (KNN) and missForest [35], among others. Another challenge arises when algorithms and criteria require a minimum amount of information for learning, as well as full knowledge of the number of rules needed to represent the data. Along the same lines, Callaghan et al. [36] highlight the importance of establishing reliable criteria for evaluating the sufficiency of the information learned and used in different algorithms, identifying this need as a current research gap.

These approaches fall into four main categories: (i) sampling-based criteria, where the expert estimates the amount of relevant information through random sampling at the beginning of the process, stopping the process when the target set by the expert system is reached; (ii) heuristic-based techniques,

 doi:10.20944/preprints202510.0359.v1

where detection occurs when a certain amount of information irrelevant to the process is detected; (iii) pragmatic criteria, which set a predefined learning time for the system, adjusted to the application case, achieving low efficiency; and (iv) new automatic detection criteria, which implement novel and complex approaches to automatically decide when to stop the learning process, based on specific metrics.

This work introduces a sufficiency criterion for determining the optimal number of rules learned by a system based on motion states, to model the overall dynamics of the system.

This criterion is based on the concept of maximum entropy applied to new rules generated over time, which is particularly useful in systems with locally stationary dynamics. The system segments the camera's field of view into areas with a high-state probability of movement, then classifies these states into disjoint and connected sets, each of which is assigned a symbolic label. Over time, the system generates sequences of motion symbols, which are ultimately organized into a grammatical structure that models the succession of states using an automated criterion [37,38]. The grammatical structure becomes inferred by taking the emitted words as positive samples of the common dynamics sampled. Then, the system represents the dynamics of movement in video scenarios, using grammatical inference to identify recurring patterns and a formal criterion of sufficiency that determines when the system has achieved learning stability.

Referring to the above paragraphs, the most relevant contributions of this research are:

- The definition of a formal criterion of information sufficiency that determines when the system has reached grammatical stability.
- The criteria exploit the properties of periodic systems, based on information criteria to learn the most significant grammar rules.
- A discrete model criteria based on right grammars to symbolically represent movement trajectories in video sequences.
- A criterion to encode a state-based system, the different trajectories through the *SEQUITUR* approach for inferencing the grammar structure in the most compact representation of movement patterns.

2. Proposed Model

This proposal describes a state-based model, where each state represents a spatial area of the camera's field of view projection. A set of adjacent and connected pixels defines each state. The states do not overlap, so that the activation of a state (indication of movement) is considered a probability function of the variability of the intensity of the pixels that compose it. States with a high probability of variability are considered active and are subject to change over time. Thus, the dynamics of an object moving through the scene are expressed as a sequence of state activations over time.

The structures of observable dynamics are characterized by a stack automaton, which enables the modeling of object movements hierarchically in the form of a proper grammar. The grammar representing the movement is constructed using the Sequitur algorithm, which infers a hierarchical structure from a sequence of discrete symbols [37]. This approach enables the modeling of dynamics using hierarchical structures, and the process of verifying whether the dynamics are associated with the model involves determining, through the grammar, whether the previously defined structures can explain the sequence of activated states.

The formal description of the proposal is expressed as follows: Let the workspace V be a grid space representing the two-dimensional projection of the camera's field of view with dimensions $n \times m$. Each position in the workspace $V_{i,j}$ represents the intensity of the pixel at position (i, j) on the image. Then, the possible number of states (partitions) of the workspace V is determined by $2^{|V|}$. In this case, a state $e \in 2^{|V|}$ is a valid state if and only if, for the state $e = \{x_1, x_2, \ldots\}$ and for every tuple (a, b) where $a, b \in e$, there is at least one connected path between both points. Then, the set of states of the system is determined by $E = \{e_1, e_2, \ldots\}$, where each state is disjoint from the others, but may or may not be adjacent to others.

 doi:10.20944/preprints202510.0359.v1

On the other hand, if there is evidence that any object occludes a state, it can be calculated using temporal contrast dynamics. That is, when there is a significant change in the intensities of the pixels that comprise it, the state is considered to be occluded and, consequently, to exhibit movement. The motion equation is defined in Equation (1).

$$M_t(e;\delta) = P(\forall(i,j) \in e) > \delta \tag{1}$$

The values of δ and the estimate of P represent the confidence/certainty of motion detection and the process for estimating probability. The function $M_t(e)$ allows us to determine whether the state e has motion or not at time t. To symbolically represent the active states, an alphabet Σ is defined. From this, a bijective subset $Id \subset (E \times \Sigma)$ is constructed, which uniquely associates each active state e with a symbol σ_i from the alphabet Σ, thus allowing the generation of symbolic sequences associated with the dynamics of the system.

Finally, the function Γ allows us to know, in terms of the alphabet Σ, the states that have a probability of movement $M_t(e)$, which is defined in $\Gamma : E \to 2^{|\Sigma|}$.

$$\Gamma_t(e) = \begin{cases} \sigma_i & \text{if } M_t(e) > \delta \text{ and } (e,\sigma_i) \in Id \\ \varnothing & \text{if } M_t(e) \leq \delta \end{cases} \tag{2}$$

The function $\Gamma_t(e)$ assigns a unique symbol $\sigma_i \in \Sigma$ to each state $e \in E$ that has a significant probability of movement. This assignment is made under the condition imposed by the detection criterion $M_t(e;\delta)$, which evaluates whether the state e has a probability of movement greater than a defined threshold δ. Consequently, the function $\Gamma_t(e)$ assigns symbols from the alphabet Σ only to those states that have been classified as active. In this way, the movement over the scene activates a set of states, which under $\Gamma_t(e)$ produce a sequence of strings emitted by the scenario. These sequences become used to build the grammar of the dynamic movement. In a current frame, more than one state becomes activated for the movement appreciated in the scene. In these cases, the strings become concatenated with only adjacent states, which avoids the use of non-connected movements, and reduces to the use of only strings generated with temporal adjacent motion evidence to infer the grammar.

Finally, discrete sequences s_i are used to infer grammars G_i, considering each emitted string as a positive sample of the grammar. These grammars enable the identification of recurring movement patterns, which are expressed through grammatical rules $N_j \in G_i$. By applying the properties of SEQUITUR in the generation of rules, it is possible to construct hierarchical representations of the observed behavior.

With this approach, it is possible to establish a criterion for learning sufficiency $\Delta f(n) = 0$, based on grammatical stability. This criterion is formalized as $\Delta f(n) = d(|R_n^*|, |R_{n-1}^*|)$, where $|R_n^*| = |N_0, N_1, \dots|$ represents the number of rules inferred for the representation of $\{s_1, s_2, \dots\}$ in the i-th sequence. This criterion indicates that when a point is reached at which the number of rules generated between sequences s_{i-1} and s_i satisfies $|R_n^*| = |R_{n-1}^*|$, the system has stabilized its knowledge and captured the main dynamics of the environment.

2.1. Motion Modeling

Within the study of computer vision-based learning methods, various techniques have been proposed for representing motion information, as discussed in [31,39], which offer interesting proposals for extracting motion patterns. One of these approaches can be represented by the following expression.

$$M_o\left(V_{i,j}^t, V_{i,j}^{t-1}\right) = V_{i,j}^t - V_{i,j}^{t-1} \tag{3}$$

where M_o represents the motion information of the objects in the scene, and $V_{i,j}$ denotes the intensity of the pixel at position (i,j) at time instant t. The difference between the intensity values at two

 doi:10.20944/preprints202510.0359.v1

consecutive time instants t and $t-1$ allows us to estimate the change associated with the motion. From this information, it is possible to obtain the trajectories of the moving objects, represented as the set $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$. The information on the movement pattern generated by an object, throughout a sequence of images at a given time t, is expressed as $\tau_i = \{x_1, x_2, \ldots\}$, where each element x_i corresponds to a spatial position occupied by the i-th object at a time t [40].

2.2. Trajectory Generation Process

In the process of generating trajectories, various techniques help increase the expressiveness of the representation of motion information, among which the binarization of trajectories [41,42], N-tuples, symbol strings [43], among others, stand out. Each has its advantages and limitations, depending on the specific case study. To provide a clearer view of the alternatives reported in the literature, Table 1 summarizes representative motion coding techniques applied in similar contexts. This comparison highlights the diversity of state definitions and coding strategies, offering a reference framework for positioning the present work within the field.

Following this analysis, our approach adopts a representation based on symbol strings [44], which offers an effective solution to the problem of trajectory expressiveness τ_i. The new representation of trajectories begins with the definition of regions belonging to the movement E and is formalized by the function $\mathcal{T}^* : \mathcal{T} \to \Sigma$, where each trajectory of an object τ_i is represented as a sequence of states $\tau_i^* = \{e_1, e_2, \ldots\}$.

Table 1. Comparison of movement coding techniques in similar context applications.

Author	Motion Coding	State Definition	Remarks
Andrea Rosani et al. [45]	Hot spots	Empirical approach	The approach focuses on activity recognition for behavioral analysis in known scenarios using context-free grammars. By incorporating both positive and negative sampling, this method enhances discrimination capabilities during the retraining process, enabling better adaptation to environmental changes.
Garcia-Huerta et al. [44]	Labeled regions	Manual	The study focuses on detecting abnormal visual events within a scenario. Motion detection is achieved through temporal differences, applied as a decay function, which identifies motion zones. These zones are subsequently segmented using the Watershed transform.
Waqas Sultani et al. [5]	Labeled regions	Fixed number	This work proposes an anomaly detection model based on deep multiple instance learning. The model utilizes a fixed number of segments from both anomalous and normal surveillance videos.
M. Daldoss et al. [32]	Hot spots	Automatic	Introduces a method for analyzing trajectories by tracking objects through a two-view camera system in surveillance videos using automatically learned context-free grammars. The model is trained on a set of sequences that define rules for different behaviors.
Hernandez-Ramirez et al. (proposed)	Labeled regions	Automatic	This work introduces an approach centered on information sufficiency, where symbolic trajectories from object–labeled region interactions are used to infer grammars. The sufficiency criterion, defined by the stabilization of production rules, ensures a meaningful representation of scene dynamics.

In the same way, the process for obtaining the finite state set $E = \{e_1, e_2, \ldots\}$ resulting from the analysis of the scene using the method proposed by Garcia et al. [44], where the author begins with a differentiation process on the information of the trajectories of the moving objects in the scene, using a decay function proposed by the author in a previous work [46], $M_{\mathcal{T}} : V \times V \to V$, to subsequently obtain the history of the displacement information using the following equation:

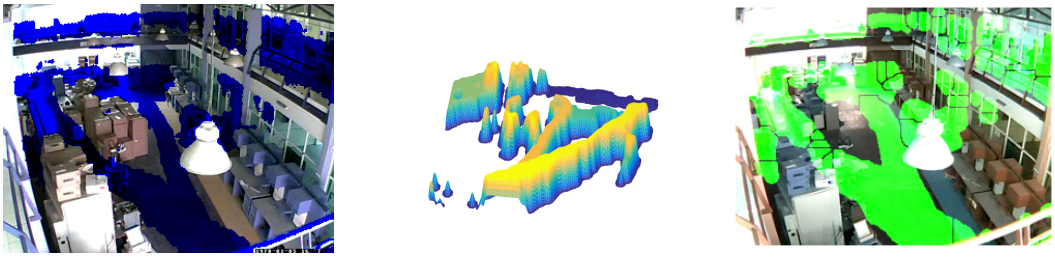
 doi:10.20944/preprints202510.0359.v1

$$M^t_{\mathcal{T}_{V_{i,j}}} = \alpha M^{t-1}_{\mathcal{T}_{V_{i,j}}} + (1-\alpha)M_o \tag{4}$$

The magnitude of the information contained in $M^t_{\mathcal{T}}$ depends on the value of the decay factor determined by α. Next, the cumulative sum defined by S_M (see Equation (5)) is calculated using the trajectory information obtained from the scene $M_{\mathcal{T}}$ to get the spatial distribution of motion membership in two dimensions.

$$S_M = \sum_{i=1}^{k} M^i_{\mathcal{T}} \tag{5}$$

To define the areas with a high probability of occurrence of movement E, the surface is segmented using movement information from the trajectories S_M. This process is obtained by applying the Watershed Transform to S_M, taking the local maxima as a reference [47–49]. This effect can be seen in Figure 1 (b), where a pseudocolor map shows the areas with the highest probability of movement in yellow, while the blue areas indicate a low probability of movement.



(a) (b) (c)

Figure 1. Process for generating path trajectories: (a) Trajectory motion information S_M, (b) Surface generated with trajectory information S_W, and (c) States mask S_R.

With the regions of E defined, now is possible to take the information of the states mask $S_R = Watershed(S_W) \cap V$, Figure 1 (c) to obtain the movement representation as trajectories S with:

$$s_i = \Gamma_t(\tau^*_i) \tag{6}$$

where the trajectory τ^*_i formed by the new set of regions E is defined as $\tau^*_i = \{\sigma_1, \sigma_2, \ldots\}$, then applying $\Gamma_t(\tau^*_i)$ a new trajectory set S is formed by the interaction of the objects with the sates mask S_R resulting from the movement information contained in τ^*_i.

2.3. Grammar Inference

With the symbolic representation of movement through a chain of symbols s_i, it is now possible to obtain right-grammars, which allow for the representation of movement patterns. A grammar is defined as a tuple $G = \{\Sigma, \mathcal{V}, R, N_0\}$ where Σ is a set of terminal symbols, $\mathcal{V}$ is the set of non-terminal symbols, R are the production rules that define how symbols can be transformed, and N_0 is defined as the initial rule [50].

Based on the above, given a path $s_1 = \{a, b, c, a, b, d, a, b, c\}$, we obtain the right grammar G_{s1} as a result of applying SEQUITUR to s_1 [51], where $G_{s1} = SEQUITUR(s_1)$. The associated alphabet is $\Sigma = \{a, b, c, d\}$, and the non-terminal symbols are $\mathcal{V} = \{N_1, N_2\}$. From this process, it can be observed that the pair of symbols (a, b) is repeated at the beginning of the sequence. Consequently, SEQUITUR detects this repetition and generates a replacement with a non-terminal symbol $N_1 \in \mathcal{V}$, finally generating the production rule $N_1 \rightarrow a, b$. This sequence is transformed into $\{N_1, c, N_1, d, N_1, c\}$, and then it is observed that the pair of symbols N_1, c is repeated once more on two occasions. This pattern is also replaced by a new non-terminal symbol $N_2 \in \mathcal{V}$, thus generating the following production rule $N_2 \rightarrow N_1, c$. Following the natural order of the process, the sequence is transformed once again,

 doi:10.20944/preprints202510.0359.v1

resulting in a compact grammar defined as $N_{0_{s_1}} = N_2, N_1, d, N_2$. This process results in the generation of the grammar G_{s_1}. This extensive process is illustrated in Table 2.

Table 2. Process for generating grammars with SEQUITUR.

Input	String	Grammar	Rules
a	$abca$	$N_{0_{s_1}} \rightarrow abca$	
b	$abcab$	$N_{0_{s_1}} \rightarrow abcab$ $N_{0_{s_1}} \rightarrow N_1 c N_1$	$N_1 \rightarrow ab$
d	$abcabd$	$N_{0_{s_1}} \rightarrow N_1 c N_1 d$	
a	$abcabda$	$N_{0_{s_1}} \rightarrow N_1 c N_1 da$	
b	$abcabdab$	$N_{0_{s_1}} \rightarrow N_1 c N_1 dab$ $N_{0_{s_1}} \rightarrow N_1 c N_1 d N_1$	
c	$abcabdabc$	$N_{0_{s_1}} \rightarrow N_1 c N_1 d N_1 c$ $N_{0_{s_1}} \rightarrow N_2 N_1 d N_2$	$N_2 \rightarrow N_1 c$

Finally, when a dictionary R^* is created with the production rules $N \in R$ generated for G_{s_1}, it is possible to reuse them to represent other trajectories with similar structures.

3. Information Sufficiency Criterion

Given the general process of generating right grammars described in Section 2, where $G_{s_i} = SEQUITUR(s_i)$, the relationships between the symbol strings s_i and their respective grammars G_{s_i} are analyzed. The objective is to understand how the algorithm's behavior in generating production rules affects the total number of rules required to represent a given sequence. In this context, the system receives an input string and produces a unique combination of rules that completely describe it.

This analysis leads to the main contribution of this work: determining the minimum amount of information sufficient in a sequence of symbols s_i to uniquely define its right-grammar G_{s_i}.

3.1. Movement Paths τ_i and Regions e_i

The previous section defined the process for generating trajectories S. These trajectories are composed of information on the movement patterns of $\mathcal{T}$, satisfying $x_k \in e$. This implies that it is likely that more than one element x_k describing the trajectory of an object τ_i belongs to the same region e. Addressing this idea, the following homologous trajectories are proposed: $\tau_1 = \{x_1, x_2, \ldots\}$ and $\tau_1^* = \{e_1, e_2, \ldots\}$, where $\{x_2, x_3\} \in e_2$. This condition validates the property that for the trajectory $\tau_1^* = \{e_1, e_2, \ldots\}$, taking the Levenshtein distance $d_L(s_i, \sigma_l)$ as the metric, it holds that $d_L(\Gamma_t(\tau_1^*), \Gamma_t(e_2)) > 1$, for $e_2 \in \tau_1^*$. This characteristic shows us the path to the contribution of this work, indicating that the information to represent the movement in both representations is equivalent, so it is possible to infer a production rule N_i that encapsulates this recurrence. Consequently, such rules are more expressive, as they not only compress the representation but also highlight structural regularities in the motion patterns.

3.2. Movement Regions and State Discovering Process

The information on the temporality of the dynamics of the objects in the scene is obtained using Equation (4). When creating a mask with the trajectory information $\mathcal{T}$, we obtain the movement surface $S_M = \{M_{\mathcal{T}}^1 \cup M_{\mathcal{T}}^2 \cup \ldots \cup M_{\mathcal{T}}^n\}$, where each S_M with the temporal information of the dynamics of the objects, given as $M_{\mathcal{T}}^T = \{M_{\mathcal{T}}^{t_1}, M_{\mathcal{T}}^{t_2}, \ldots, M_{\mathcal{T}}^{t_n}\}$, is formed with the trajectory information τ_i, therefore there is at least one pair of trajectories τ_i and $\tau_{i+\rho}$ for which $\mathcal{A}(M^{t_i, t_{i+\rho}}) = \mathcal{A}(M^{t_i})$, where $\mathcal{A}$ represents the area operator, indicating that the trajectory $\tau_i = \tau_{i+\rho}$ is therefore a repeated trajectory. This characteristic, when extrapolated to the relationship between the mask with the temporal information of the dynamics of objects S_M and the temporal information of the dynamics of objects $M_{\mathcal{T}}$, validates

 doi:10.20944/preprints202510.0359.v1

that if two trajectories, at instants i and $i+\rho$, are equal, $\tau_i = \tau_{i+\rho}$, then there also exists a two-dimensional surface S_M, for which $\mathcal{A}(S_M(M^{t_i})) = \mathcal{A}(S_M(M^{t_i} \cup M^{t_i+\rho}))$ and therefore, to obtain the regions E, it holds that $Watershed(S_M(M^{t_i})) = Watershed(S_M(M^{t_i} \cup M^{t_i+\rho}))$. This suggests that there are trajectories τ_i for which no new regions $e_l \in E$ are being generated. In the absence of new information to represent the new set of trajectories s_i, it is necessary to define a criterion that indicates when the information from the new trajectories is no longer contributing to the formation of more regions e_l, in order to systematize this task.

3.3. Automatic Rules Discover

In the process, it is necessary to have a way of identifying the moment when the trajectories τ_i obtained to form the states e_l are not providing new information. To solve this problem, the method described in the diagram in Figure 2, which presents a methodology for defining a criterion of information sufficiency ρ. This criterion is directly associated with the number of new production rules $N_{s_i} \in R \in G_{s_i}$, which are generated when modeling the dynamics s_i of the objects in the scene. This sufficiency criterion ρ helps to automatically determine when no new information is being found, i.e., when $\Delta f(t) \approx 0$, At this point, the production rules N_{s_i} are beginning to have a high degree of repetitiveness, indicating that the modeled trajectories G_{s_i} no longer contribute significant new information.

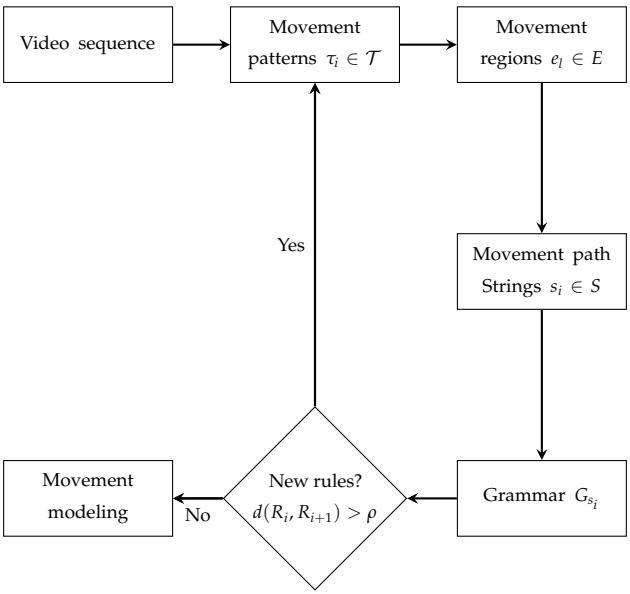


Figure 2. Automatic selection process for rules generated.

In order to determine whether there is new information in the trajectories s_i obtained, a dictionary $R^* : R^n \to R^m$ is defined. The process can be seen in Figure 2, where for each string s_i a grammar G_{s_i} is obtained that generates a set of rules $N_i \in R^*$ that satisfy:

$$R^* = \begin{cases} R^*_{i+1} & \text{si } d_L(R^*, N_i) > 0 \\ R^*_i & \text{Other case} \end{cases} \tag{7}$$

Equation (7) defines the similarity criterion for classifying new rules to be added to dictionary R^*. In relation to the above, let $f : \mathbb{Z} \to \mathbb{Z}$ be a discrete function that counts the total number of production rules in R^* accumulated at a given moment in time t.

$$f^t(n) = |R^*_i| \tag{8}$$

At the beginning of the motion modeling process using Grammars G, it is true that $f(n)^t \neq f(n)^{t+1}$, which reflects the presence of a high generation of new production rules in the dictio-

 doi:10.20944/preprints202510.0359.v1

nary R^*. Over time, part of the information in the trajectories begin to repeat themselves, that is, $d_L(\Gamma_t(\tau_1^*), \Gamma_t(e_l)) > 1$, which is directly associated with the number of new production rules incorporated into the dictionary $|R_i^*| = |R_{i+1}^*|$. This condition translates into $f(n) = c$. When there is a stable value for c, it is possible to identify the moments when $\Delta f(n) = 0$, which exposes a state of pseudo-stability. This state does not tend to be permanent, but it repeats itself for increasingly longer periods of t, which demonstrates the possibility of making approximations by referring to the condition $\Delta f(n) = 0$. However, there is a significant limitation, which consists of the variable stability of the criterion, as it does not remain completely constant.

To do this, $f : \mathbb{R} \to \mathbb{R}$ is performed with a function that makes a continuous approximation of the discrete values $f(n)$, applying the logarithmic regression method [52,53]:

$$y = a \ln x + b \tag{9}$$

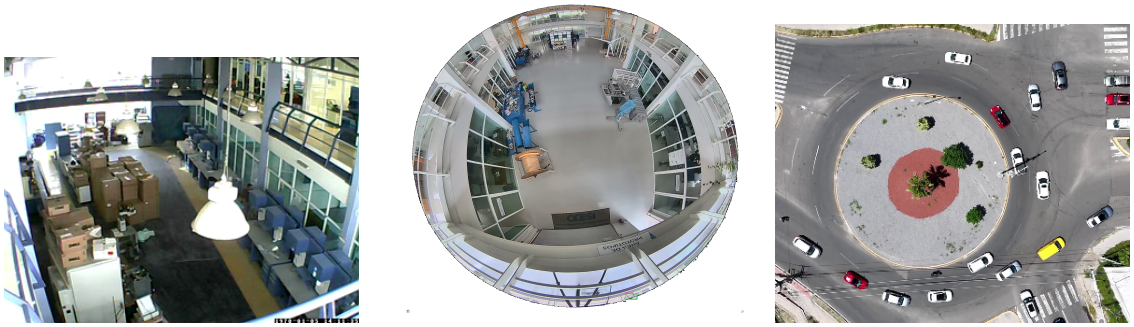
where the values of a and b in Equation (9) are obtained using the least squares technique [54], thereby determining the curve C, which shows the variation of $|R^*|$. At this point, the central contribution of this work becomes clear, namely the criterion ρ for information sufficiency, which can now be represented as:

$$y' = \rho \tag{10}$$

Due to the nature of the logarithmic function, it is not possible to define the sufficiency criterion as $y' = 0$, as this would cause an overfitting problem $(t \to \infty)$. The recommended value for the information sufficiency criterion is defined as $\rho = 1$ for Equation (10). Given that the information entered into the system is N_i, then when $\rho \geq 1$, we have $|R_i^*| \neq |R_{i+1}^*|$, otherwise $|R_i^*|1 \geq p_{min}$.

4. Experimental Analysis and Results

In this section, the proposal made in Sections 2 and 3 is validated, obtaining video sequence information corresponding to three test scenarios. The first scenario corresponds to a PTZ camera capturing a side view of the analyzed scenario (scenario 1), while the second corresponds to a static 360° fisheye camera with a top view of the scenario (scenario 2). Finally, the third scenario (scenario 3) was obtained from an aerial shot using a wide-angle camera. These scenarios can be observed in Figure 3.



(a) (b) (c)

Figure 3. Scenarios used to validate the proposal: (a) $\mathcal{E}_1$; (b) $\mathcal{E}_2$ and (c) $\mathcal{E}_3$.

4.1. Experimental Process

To validate the proposal, we propose a process for automatically obtaining states e_l of motion, which are directly related to $\Delta f(n)$. The proposed method consists of the following stages:

1. *Obtaining trajectories.* The trajectories of the movement observed in the scene represented by the set $\mathcal{T}$ are obtained during the scene modeling process. This set contains multiple trajectories detected and enumerated incrementally, defined as $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$. Each trajectory t_i is composed of a sequence of positions $\tau_i = \{x_1, x_2, \ldots\}$, which describe the movement pattern associated with the

 doi:10.20944/preprints202510.0359.v1

intensity of pixel (i, j) at position $V_{i,j}$ corresponding to the movement captured in the scene over time t.

2. *Movement states definition.* With the information on the movement of objects contained in $\mathcal{T}$, it is necessary to understand the dynamics of the objects in the scene over time. This temporal representation is achieved by the function $M_{\mathcal{T}}(\mathcal{T})$, which relates the movement information of the trajectories to the membership of each pixel (i, j) in the movement. By analyzing the belonging to the movement using $M_{\mathcal{T}}(\mathcal{T}) = \{M_{\mathcal{T}}(\tau_1), M_T(\tau_2), \ldots, M_T(\tau_n)\}$, an accumulated surface S_M is formed that contains the integrated information of the movement patterns. By applying the *Watershed* algorithm to S_M, a two-dimensional representation of the regions with the highest probability of belonging to the movement is obtained, which are defined in the set of states E.
3. *Symbol strings.* At this stage, once the states E have been defined, the state mask $S_w \subset (E \times \Sigma)$ is constructed, which relates each state $e_l \in E$ to a symbol $\sigma_i \in \Sigma$. The relationship between the movement positions x_k and the states e_l forms the state trajectories t_i^*. Applying $\Gamma(t_i^*)$ yields the symbolic sequences s_i.
4. *Sufficiency of information.* The movement sequences s_i are used to infer grammars by means of $G_{s_i} = SEQUITUR(s_i)$. This process generates a dictionary of production rules R^*, where the change in the cardinality of the set of production rules is taken as a reference to define the criterion $\Delta f(n) = \rho$ to define when the system has learned enough, $f(n)$ shows the increase in the number of rules. Naturally, when analyzing the change in $f(n)$, we see that when $\Delta f(n)$ meets the condition $\Delta f(n) = 0$, this means that $\tau_i = \tau_{i+1}$, therefore, no new production rule is being generated in the dictionary. This characteristic of the behavior of the function $f(n)$ is a clear indication that a point of sufficiency of acquired information is being reached.
5. *Dynamics modeling.* Once sufficient information has been defined to model the scene, activities are detected by applying SEQUITUR to the movement sequences obtained.

4.2. Results

Information from the video sequences corresponding to the scenarios Figure 3, described below, is used. $\mathcal{E}_1$ shows the interior of a laboratory with low light variability through a fixed camera with a resolution of 240×320 pixels; $\mathcal{E}_2$ captures the scene using a fish-eye camera under controlled lighting conditions with a resolution of 1920×1920 pixels; and the third scenario $\mathcal{E}_3$, which acquires information on traffic flow at a roundabout using a drone with a resolution of 2160×3840 pixels, where there is no control over the lighting.

Following the methodology proposed in the previous section Figure 2, the videos are analyzed in order to extract the movement patterns. With the information on the moving objects M_o when applying Equation (3), it is possible to obtain the movement patterns $\tau_i = \{x_1, x_2, \ldots\}$, to determine the temporality of the movement $M_{\mathcal{T}}$, allowing the movement membership of each pixel $V_{i,j}$ to be visualized, as described in Equation (4).

Once the information on belonging to the movement $M_{\mathcal{T}}$ has been obtained, it is easy to observe how the movement zones in the scene change with respect to time $M_{\mathcal{T}}^t$, which, when applying S_M, shows that as time passes, the moving objects cover most of the areas of the scenes, a process that can be observed in Figure 5.

The information on the movement positions $x_i \in V$ stored by each $\tau_i \in \mathcal{T}$ can be seen in the movement areas in Figure 4. It can be inferred that the information obtained in S_M is related to a high degree of information about the movement of objects in the scene $M_{\mathcal{T}}(\mathcal{T})$ Figure 5.

With the temporal representation of movement using the function $M_{\mathcal{T}}^t(\mathcal{T}_{\mathcal{E}})$, it is possible to observe how the information on movement patterns is distributed throughout the scene with $S_M = \{M_{\mathcal{T}}^1 \cup M_{\mathcal{T}}^2 \cup \ldots \cup M_{\mathcal{T}}^5\}$ for $t = 5$ minutes. Al aplicar *Watershed* a la superficie acumulada S_M se obtienen los estados E que representan las regiones donde existe mas probabilidad de pertenecia al movimiento, este proceso puede observarse en la siguiente Figure 6.

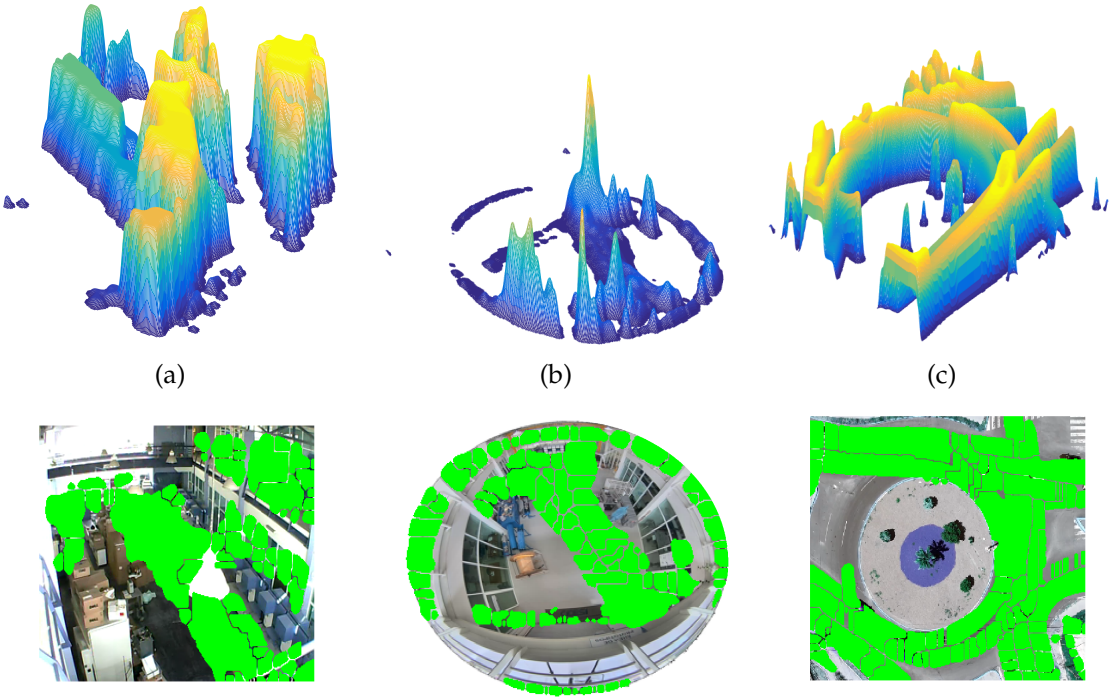
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 October 2025 doi:10.20944/preprints202510.0359.v1

(a) (b) (c)

Figure 4. Obtaining temporality of movement $M_{\mathcal{T}}$: (a) Detection of movement x_k; (b) Trajectory t_1; and (c) Temporality of movement of $M_{\mathcal{T}}(\tau_1)$.

(a) (b) (c)

Figure 5. Trajectories $M_{\mathcal{T}}(\mathcal{T})$ for the three scenarios, obtained at different time intervals t. (a),(b) and (c) Movement temporality for $M^t_{\mathcal{T}}(\mathcal{T}_{\mathcal{E}})$ for t equal to 5 minutes.



(a) (b) (c)

(d) (e) (f)

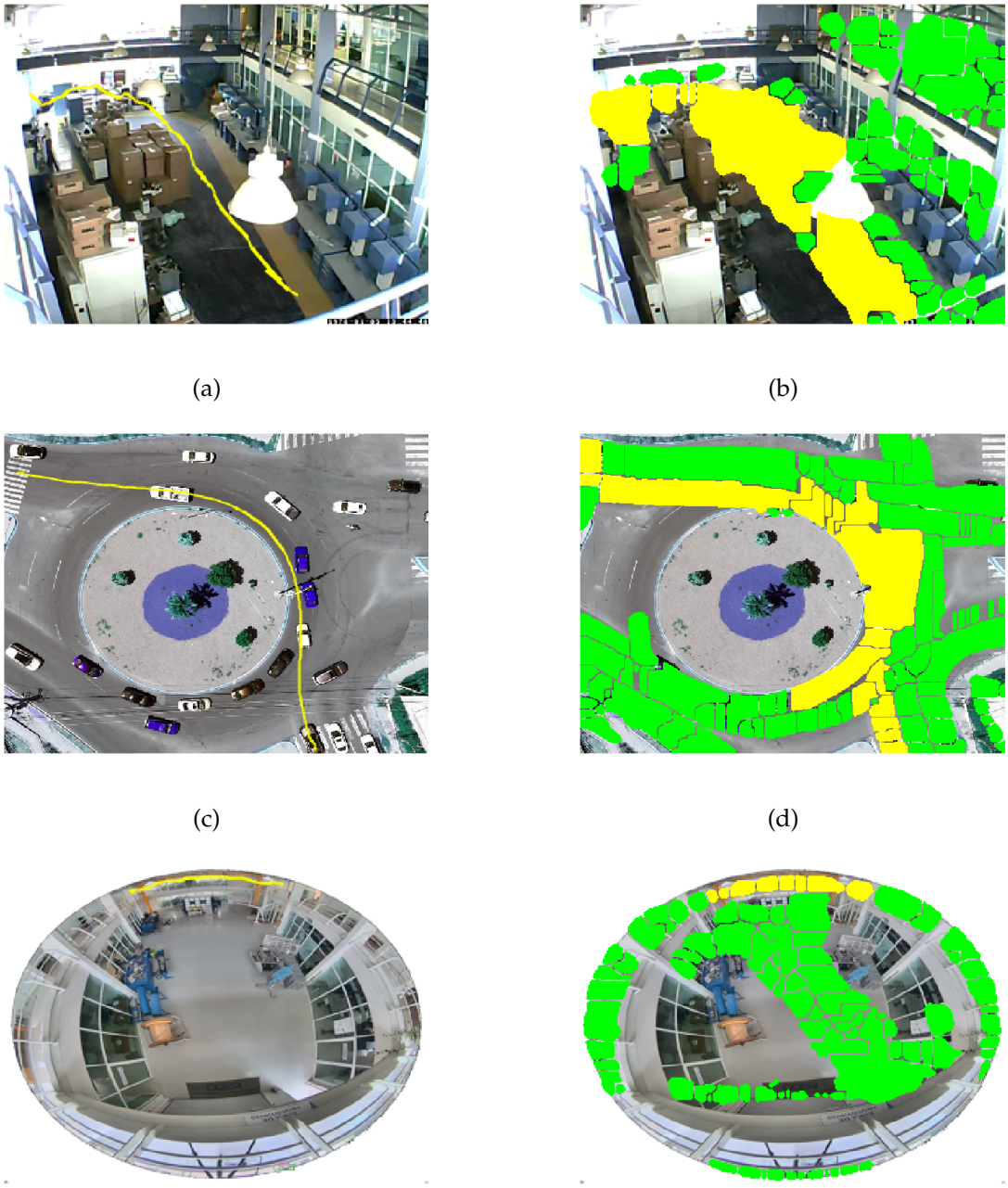
Figure 6. Definition of states E for the three scenarios with $t = 5$ minutes: (a-c) Surface with movement pattern information S_M in scenarios $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$; (d-f) State mask S_w for the three scenarios.

After defining each state within the scenarios and associating it with the corresponding trajectories, the trajectories $\tau_i^* \in \mathcal{T}_i^* \subset E_{\mathcal{E}}$ are obtained. From this, it is possible to apply $\Gamma(\tau_i^*)$ to obtain the symbolic representation s_i of the movement patterns. In this process, each state e_i is assigned a symbol σ. For this work, the symbols used are $\sigma \in \mathbb{N}$. The trajectory $\tau_1 = \{x_1, x_2, \ldots, x_k\}$ in Figure 7 (a) can

 doi:10.20944/preprints202510.0359.v1

now be represented as $s_1 = \{\sigma_1, \sigma_2, \ldots, \sigma_k\}$. This representation allows us to infer a grammar $G_{s_{1_{\mathcal{E}_1}}}$ for this symbolic representation.

The relationship between the growth in the number of rules generated for the grammars inferred on the trajectories obtained in 5 minutes is shown in Figure 8. The incremental behavior in the number of production rules generated when inferring the grammars for the total trajectories obtained in the time defined as 5 minutes. Applying the same principle to Figure 7 (f), with the grammar $G_{s_{1_{\mathcal{E}_3}}}$, we observe that 14 new rules were generated corresponding to $|R| = 14$.



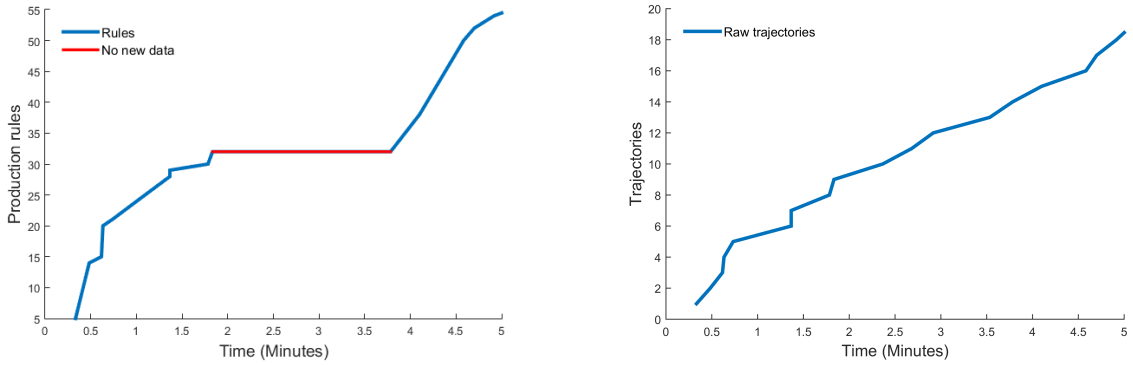
(a) (b) (c) (d) (e) (f)

Figure 7. Trajectories in their representations τ_1 and s_1: (a) Trajectory τ_1 for scenario $\mathcal{E}_1$; (b) Trajectory s_1 for scenario $\mathcal{E}_1$; (c) Trajectory τ_1 for scenario $\mathcal{E}_2$; (d) Trajectory s_1 for scenario $\mathcal{E}_2$; (e) Trajectory τ_1 for scenario $\mathcal{E}_3$; and (f) Trajectory s_1 for scenario $\mathcal{E}_3$.

 doi:10.20944/preprints202510.0359.v1

The relationship between trajectories and grammars can be seen in Figure 8, which shows the behavior of the production rules R for $\mathcal{E}_1$. The number of trajectories obtained over time is not related to the number of production rules R^* generated to model the different trajectories s_i.

This property allows us to begin the information sufficiency process, which consists of performing a classification described in Equation (7), where the objective of the process is to discriminate those trajectories that are not generating new information. Each rule N_i that is added to the dictionary is selected by applying a similarity criterion $d_L(R^*, N_i) > 0$ or, when a trajectory does not generate new information, $d_L(R^*, N_i) = 0$ is satisfied, with the information from the trajectories that generate new rules that can be included in the dictionary R^* it is possible to form a discrete function $f(n)$, which stores the cardinality of the dictionary with respect to time t. When graphing this discrete function $f(n)$, what can be observed is that as the information analysis time increases, the production rules $N_i \in R^*$ comply more with $R_i^* = R_{i+1}^*$, this can be seen in Figure 8.



(a) (b)

Figure 8. Relationship between the number of production rules vs. trajectories with $t = 5$ minutes: (a) Production rules $|R|$ for scenario $\mathcal{E}_1$ and (b) behavior of the number of trajectories over time t for the first scenario.

The parts in red show the moments in time when no new information was generated, despite constantly obtaining movement information. This means that one trajectory s_i shares information with another trajectory s_j.

For this proposal, repetitiveness is an indicator that shows that the new movement information $V_{i,j}$ obtained from the scenario begins to lose significance $\Delta f(n) = 0$. Up to this point, it is difficult to define a criterion for information sufficiency based on the number of repeated rules in the dictionary R^* over time.

This leads to focusing efforts on approximating the discrete function $f(n)$ with a continuous function $f(t)$. This approximation is possible by calculating the values a and b and substituting them into Equation (9), in order to obtain the continuous values of $f(t)$ that best describe the growth curve of the production rules obtained over time in each of the scenarios. This process can be seen in Figure 10.

Taking advantage of the characteristics of the continuous function $f(t)$, we focus on analyzing how this function changes at points where there is high repetitiveness of production rules $N_i \in R^*$. Now it is possible to apply a criterion of information sufficiency based on the change in the approximate function Equation (10), which is defined as $f(t)' = rho$, where the function $f(t)'$ provides information on the growth of the production rules obtained over time. The minimum value of information desired to define that the new movement information obtained is significant is when at least one new production rule is being obtained. Then this criterion ρ is passed to the definition of information sufficiency with value $\rho = 1$.

The sufficiency criterion $f(t)' = 1$ is applied in the study scenarios in order to obtain the automatic definition of the movement information time necessary to determine that the information analyzed has been sufficient. For scenario 1, the time t associated with this criterion is $t_1 = 98.55$ minutes, and for scenario 2, the time $t_2 = 189.4$ minutes (see Figure 10).

 doi:10.20944/preprints202510.0359.v1

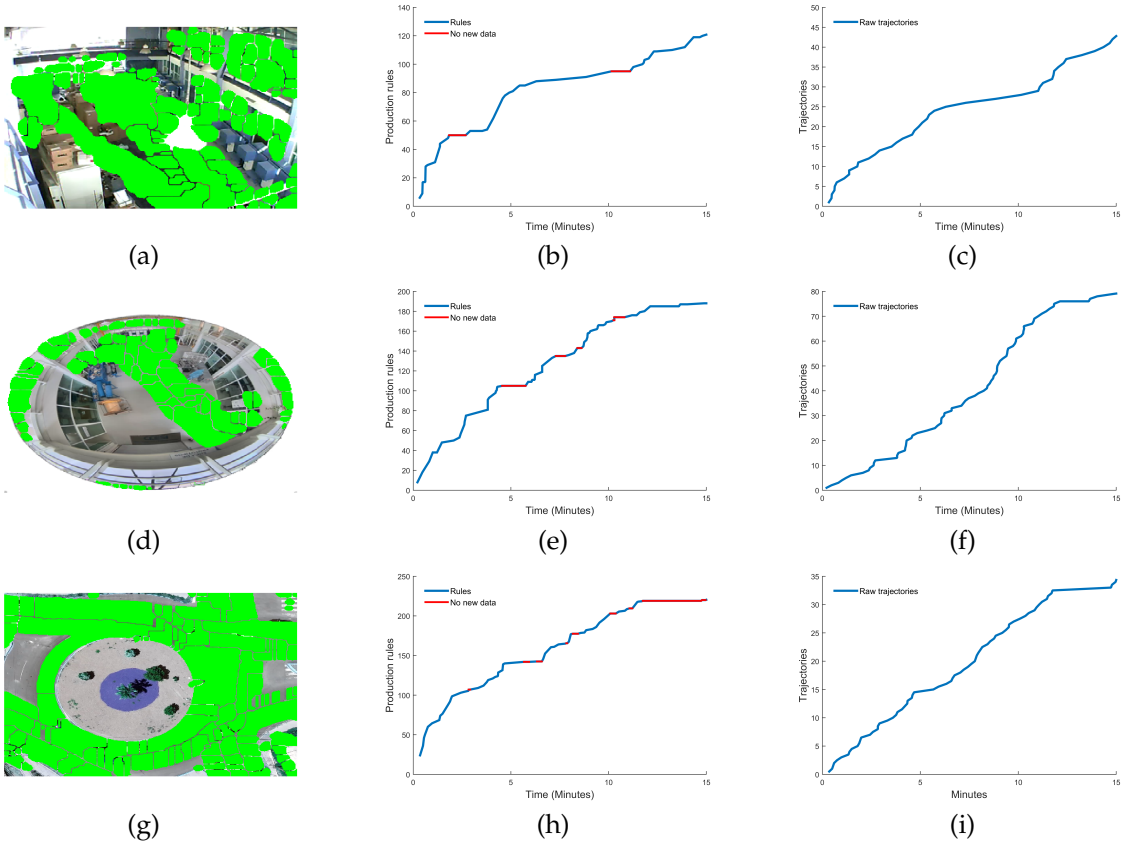


Figure 9. Relationship between the behavior of states $E_{\mathcal{E}}$, production rules, and trajectories, corresponding to $t = 15$ minutes: (a, d, g) states $E_{\mathcal{E}}$; (b, e, h) production rules R^*; and (c, f, i) generated trajectories S.

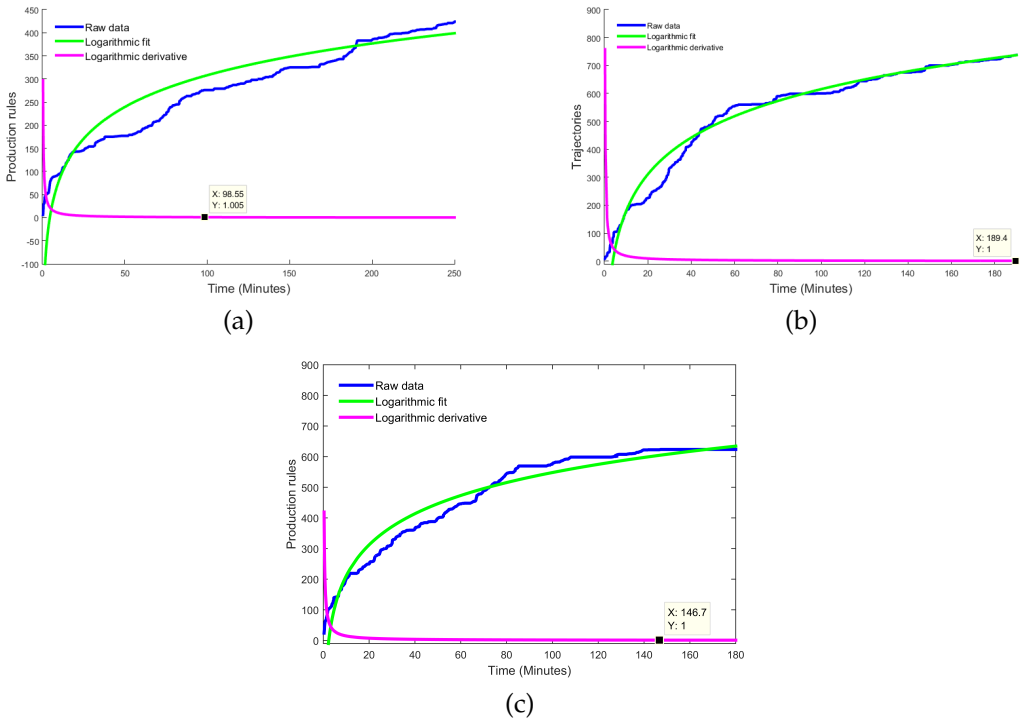


Figure 10. Time required for sufficient movement information $f(t)' = 1$: (a) Scenario 1 $t_1 = 98.55$ minutes and $f(98.55)' = 1$; (b) Scenario 2 $t_2 = 189.4$ minutes and $f(189.4)' = 1$; (c) Scenario 3 $t_3 = 146.7$ minutes and $f(146.7)' = 1$.

 doi:10.20944/preprints202510.0359.v1

With the growth information in the $R^*_{\mathcal{E}}$ dictionaries corresponding to the analyzed scenarios, this is directly associated with the time in each scenario. For the sufficiency of information, work is being done to generate more expressive trajectories, taking the information from the accumulated movement patterns $S_{M_{98.55}}$ and $S_{M_{189.4}}$ as shown in Figure 11.

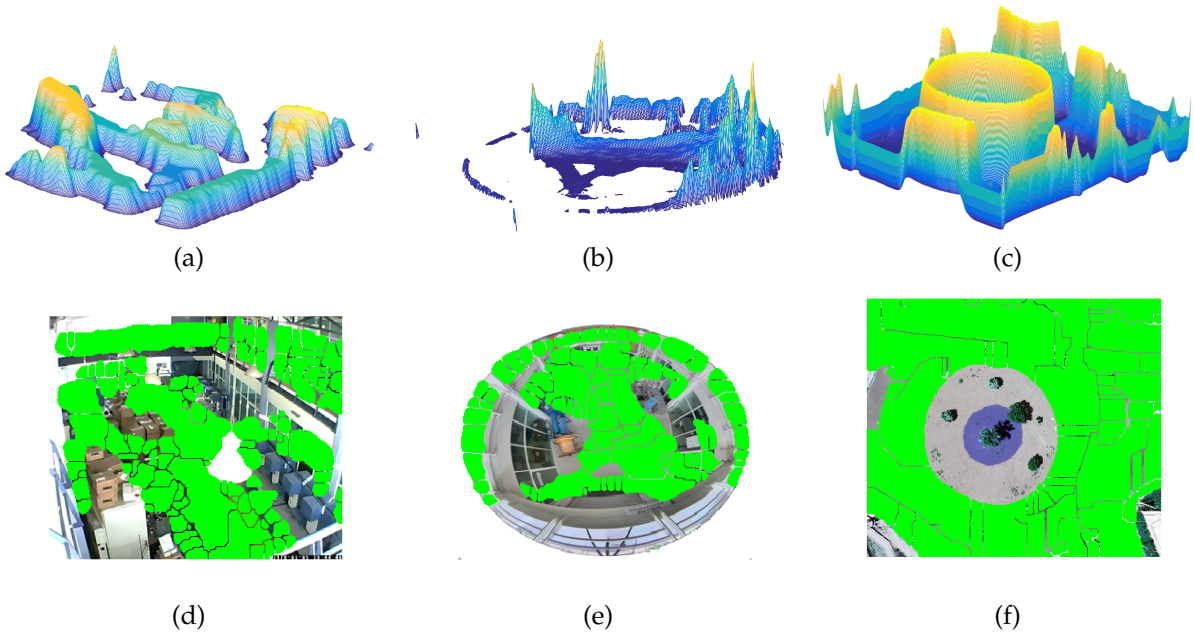


Figure 11. Membership in the movement with (a) Movement information for $S_{M_{98.55}}$; (b) for $S_{M_{189}}$; (c) for $S_{M_{67}}$; (d) Surface generated with $S_{M_{98.55}}$ and (e) $S_{M_{189.4}}$ and (f) $S_{M_{67}}$.

The information from $S_{M_{98.55}}$ is used to form a new representation of the movement of type s_i, with more expressive trajectories composed of the states e_k of the scene where the movement is detected. With this information, it is possible to model the scene for activity detection. The information needed to train the model was defined by the criterion $f(t)' = 1$. In turn, the rules learned by the model and stored in the dictionary R^* are validated to infer grammars and validate them against a new set of trajectories, thus determining the effectiveness of the model in each scenario.

The effectiveness of each model is presented in Table 3, where the results obtained in the three scenarios evaluated are compared. It can be seen that efficiency varies depending on the particular morphology of each scenario. This validation supports the proposal put forward in this work, demonstrating that the criterion of information sufficiency is capable of adapting to different contexts while maintaining consistent levels of performance.

Table 3. Results of trajectory experimentation in three different scenarios.

Scenarios	Training time	Total trajectories $\lvert S_{\mathcal{E}} \rvert$	Dictionary rules $\lvert R^*_{\mathcal{E}} \rvert$	Accuracy (%)
(1) PTZ Camera	98.55 minutes	440	250	84.13 %
(2) Fisheye 360° Camera	189.4 minutes	1022	700	83.56 %
(3) Wide angle Camera	146.7 minutes	319	622	95.92 %

5. Discussion and Conclusions

Learning algorithms are widely used today; however, like any process, they have areas for improvement and opportunities for optimization. One of the most relevant and widely studied areas today is the need for large volumes of information to train them properly. Obtaining this data can be complex and costly, both in terms of time and computational resources. In this regard, it is essential to

 doi:10.20944/preprints202510.0359.v1

have proposals such as the one presented in this paper, which introduces a new criterion for defining the sufficiency of information required in the learning of this type of algorithm. This section discusses the results obtained with the information sufficiency criterion, emphasizing the objectives achieved, the benefits obtained, and the limitations identified.

The application of the proposed method in the three scenarios evaluated proved to be an efficient tool for establishing a new criterion of information sufficiency, taking advantage of the relationship between the grammatical rules generated by the model and the dynamics of the objects in the analyzed scene. The Section 4, dedicated to experimentation, presents the results of trajectory modeling using grammars in three different scenarios. These results, summarized in Table 3, show the analysis of the model in each case.

In the first scenario, captured with a PTZ camera, the model indicated that a training time of 98.55 minutes was sufficient for the generated data to achieve a high degree of repeatability, thus allowing the training time to be automatically determined. The resulting rule set contained $R^* = 250$ rules, achieving an accuracy of 84.13% when validated with new trajectories. In the second scenario, recorded with a 360° fisheye camera, the model determined a training time of 189.4 minutes, resulting in a rule set of $R^* = 700$ rules and an accuracy of 83.56%. Finally, in the third scenario, captured with a wide-angle camera, the estimated training time was 146.7 minutes, with a rule set of $R^* = 622$ rules and a validation accuracy of 95.92%. These results can be seen in more detail in Table 3.

Regarding the limitations of this approach, it should be noted that the method assumes that objects move freely within the scene and that, given sufficient time, all areas will be covered by motion data. This assumption may not hold true in scenarios where movement is restricted or the areas of movement are limited.

Furthermore, the results obtained in the analyzed scenarios may vary when applied to different environments. The intrinsic parameters of the model are not fully generalized, which could affect the method's effectiveness in contexts with different characteristics.

In conclusion, this work presents an effective approach for defining a criterion for information sufficiency. The main advantage lies in representing trajectories as sequences of symbols s_i, which allows us to infer a grammar to model the dynamics of the objects. This results in rules where, if two trajectories exhibit similarities, it means that at least one production rule applies to a portion of both trajectories, thus significantly reducing the amount of information R_i needed to represent the movement. In this way, the trajectory information G_{s_i} becomes more expressive and suitable for modeling and machine learning purposes.

5.1. Further Works

The information sufficiency criterion developed in this work opens up a wide range of avenues for future research. One initial direction is the application of this criterion to the modeling of stochastic processes, particularly to Hidden Markov Models (HMM). In this context, the regions e_i that form the trajectories s_i can be directly mapped to the visible states v_k of an HMM. This correspondence offers a significant advantage: it ensures that the number of visible states is determined by the information sufficiency, thus avoiding both overfitting and under-representation of the underlying dynamics. Furthermore, the grammatical structure can serve as a guide for estimating the transition and emission probabilities, providing an additional layer of regularization during HMM training.

Finally, this approach can be extended to other domains where symbolic representation of data is useful, such as multivariate time series analysis, pattern recognition in biological sequences, or predicting behavior in complex environments. In all these cases, the grammar induced from the data, along with the concept of information sufficiency, could contribute to improving the generalization ability of models and reducing training costs.

Author Contributions: Conceptualization, H.H.-R. and H.J.-H.; Formal analysis, A.-M.H.-N.,D.C.-E. and H.J.-H.; Methodology, H.H.-R., A.-M.H.-N. and H.J.-H.; Software, H.J.-H.; Supervision, H.H-R., D.C.-E. and H.J.-H.;

 doi:10.20944/preprints202510.0359.v1

Writing—original draft, J.-L.P.-R. and A.-M.H.-N.; Writing—review and editing, H.J.-H., J.-L.P.-R. and H.H.-R.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data availability in this study is on request from the corresponding author.

Acknowledgments: We wish to give thanks to CIDESI, CONAHCYT, and the Mexican Artificial Intelligence Alliance, with the FORDECYT Project 296737 *Consorcio en Inteligencia Artificial* which provided the student scholarships and CIICCTE (Centro de Investigación e Innovación en Ciencias de la Computación y Tecnología Educativa) which provided technical and infrastructure support.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
HMM	Hidden Markov Models
KNN	K-nearest neighbors
PTZ	Pan, Tilt, and Zoom

References

1. Fedorov, A.; Nikolskaia, K.; Ivanov, S.; Shepelev, V.; Minbaleev, A. Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data* **2019**, *6*, 1–15. https://doi.org/10.1186/s40537-019-0234-z.
2. Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S. A review of video surveillance systems. *Journal of Visual Communication and Image Representation* **2021**, *77*, 103116. https://doi.org/10.1016/j.jvcir.2021.103116.
3. Ess, A.; Leibe, B.; Schindler, K.; Van Gool, L. A mobile vision system for robust multi-person tracking. In Proceedings of the 2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008, pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587581.
4. Mabrouk, A.B.; Zagrouba, E. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications* **2018**, *91*, 480–491. https://doi.org/10.1016/j.eswa.2017.09.029.
5. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6479–6488. https://doi.org/10.1109/CVPR.2018.00678.
6. Wibowo, M.E.; Ashari, A.; Putra, M.P.K.; et al. Improvement of Deep Learning-based Human Detection using Dynamic Thresholding for Intelligent Surveillance System. *International Journal of Advanced Computer Science and Applications* **2021**, *12*. https://doi.org/10.14569/IJACSA.2021.0121053.
7. Sreenu, G.; Durai, S. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. *Journal of Big Data* **2019**, *6*, 1–27. https://doi.org/10.1186/s40537-019-0212-5.
8. Luck, S.J.; Stewart, A.X.; Simmons, A.M.; Rhemtulla, M. Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology* **2021**, *58*, e13793. https://doi.org/10.1111/psyp.13793.
9. Roh, Y.; Heo, G.; Whang, S.E. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering* **2019**, *33*, 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162.
10. Rijali, A. Analisis data kualitatif. *Alhadharah: Jurnal Ilmu Dakwah* **2018**, *17*, 81–95. https://doi.org/10.18592/alhadharah.v17i33.2374.
11. Denes, G.; Jindal, A.; Mikhailiuk, A.; Mantiuk, R.K. A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution. *ACM Transactions on Graphics (TOG)* **2020**, *39*, 133–1. https://doi.org/10.1145/3386569.3392411.
12. He, Y.; Wei, X.; Hong, X.; Shi, W.; Gong, Y. Multi-target multi-camera tracking by tracklet-to-target assignment. *IEEE Transactions on Image Processing* **2020**, *29*, 5191–5205. https://doi.org/10.1109/TIP.2020.2980070.
13. Gilroy, S.; Jones, E.; Glavin, M. Overcoming occlusion in the automotive environment—A review. *IEEE Transactions on Intelligent Transportation Systems* **2019**, *22*, 23–35. https://doi.org/10.1109/TITS.2019.2956813.

 doi:10.20944/preprints202510.0359.v1

14. Wang, A.; Sun, Y.; Kortylewski, A.; Yuille, A.L. Robust object detection under occlusion with context-aware compositionalnets. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 12645–12654. https://doi.org/10.1109/CVPR42600.2020.01266.
15. Li, H.; Chen, G.; Li, G.; Yu, Y. Motion guided attention for video salient object detection. In Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 7274–7283. https://doi.org/10.1109/ICCV.2019.00737.
16. Zhu, Y.; Newsam, S. Motion-aware feature for improved video anomaly detection. *arXiv preprint arXiv:1907.10211* **2019**. https://doi.org/10.48550/arXiv.1907.10211.
17. Griffin, B.A.; Corso, J.J. Depth from camera motion and object detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1397–1406. https://doi.org/10.1109/CVPR46437.2021.00145.
18. Jain, J.; Singh, A.; Orlov, N.; Huang, Z.; Li, J.; Walton, S.; Shi, H. Semask: Semantically masked transformers for semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 752–761. https://doi.org/10.1109/ICCVW60793.2023.00083.
19. Sun, Z.W.; Hua, Z.X.; Li, H.C.; Zhong, H.Y. Flying Bird Object Detection Algorithm in Surveillance Video Based on Motion Information. *IEEE Transactions on Instrumentation and Measurement* **2024**, *73*, 1–15. https://doi.org/10.1109/TIM.2023.3334348.
20. Dong, G.; Zhao, C.; Pan, X.; Basu, A. Learning Temporal Distribution and Spatial Correlation Toward Universal Moving Object Segmentation. *IEEE Transactions on Image Processing* **2024**, *33*, 2447–2461, [2024 Mar 29]. https://doi.org/10.1109/TIP.2024.3378473.
21. Zhang, T.; Huang, B.; Wang, Y. Object-occluded human shape and pose estimation from a single color image. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 7376–7385. https://doi.org/10.1109/CVPR42600.2020.00740.
22. Liu, D.; Li, Y.; Lin, J.; Li, H.; Wu, F. Deep learning-based video coding: A review and a case study. *ACM Computing Surveys (CSUR)* **2020**, *53*, 1–35. https://doi.org/10.1145/3368405.
23. Vu, H.N.; Nguyen, M.H.; Pham, C. Masked face recognition with convolutional neural networks and local binary patterns. *Applied Intelligence* **2022**, *52*, 5497–5512. https://doi.org/10.1007/s10489-021-02728-1.
24. Meinhardt, T.; Kirillov, A.; Leal-Taixe, L.; Feichtenhofer, C. Trackformer: Multi-object tracking with transformers. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 8844–8854. https://doi.org/10.1109/CVPR52688.2022.00864.
25. Kumar, S.; Yadav, J.S. Video object extraction and its tracking using background subtraction in complex environments. *Perspectives in Science* **2016**, *8*, 317–322. Recent Trends in Engineering and Material Sciences, https://doi.org/https://doi.org/10.1016/j.pisc.2016.04.064.
26. Pandey, S.; Jain, P.; Patel, P. Video Background Subtraction Algorithms for Object Tracking. In Proceedings of the 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), 2022, pp. 464–469. https://doi.org/10.1109/ICICICT54557.2022.9917717.
27. Chen, X.; Williams, B.M.; Vallabhaneni, S.R.; Czanner, G.; Williams, R.; Zheng, Y. Learning active contour models for medical image segmentation. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 11632–11640. https://doi.org/10.1109/CVPR.2019.01190.
28. Li, L.; Zhou, T.; Wang, W.; Li, J.; Yang, Y. Deep hierarchical semantic segmentation. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1246–1257. https://doi.org/10.1109/CVPR52688.2022.00131.
29. Tao, A.; Sapra, K.; Catanzaro, B. Hierarchical multi-scale attention for semantic segmentation. *arXiv preprint arXiv:2005.10821* **2020**. https://doi.org/10.48550/arXiv.2005.10821.
30. Wu, L.; Wang, Q.; Jian, M.; Zhao, Y.Q.B. A Comprehensive Review of Group Activity Recognition in Videos. *International Journal of Automation and Computing* **2021**, p. 17. https://doi.org/10.1007/s11633-020-1258-8.
31. A. Yilmaz, O.J.; Shah, M. Object tracking: a Survey. *ACM Computing surveys* **2006**, *38*, 13. https://doi.org/10.1145/1177352.1177355.
32. Daldoss, M.; Piotto, N.; Conci, N.; De Natale, F.G. Learning and matching human activities using regular expressions. In Proceedings of the 2010 IEEE International Conference on Image Processing. IEEE, 2010, pp. 4681–4684. https://doi.org/10.1109/ICIP.2010.5653507.
33. Li, D.; Wu, M. Pattern recognition receptors in health and diseases. *Signal transduction and targeted therapy* **2021**, *6*, 291. https://doi.org/10.1038/s41392-021-00687-0.
34. Braga-Neto, U. *Fundamentals of pattern recognition and machine learning*; Springer, 2020. https://doi.org/10.1007/978-3-031-60950-3.

 doi:10.20944/preprints202510.0359.v1

35. Emmanuel, T.; Maupong, T.; Mpoeleng, D.; Semong, T.; Mphago, B.; Tabona, O. A survey on missing data in machine learning. *Journal of Big data* **2021**, *8*, 1–37. https://doi.org/10.1186/s40537-021-00516-9.
36. Callaghan, M.W.; Müller-Hansen, F. Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews* **2020**, *9*, 1–14. https://doi.org/10.1186/s13643-020-01521-4.
37. Nevill-Manning, C.G.; Witten, I.H. Identifying hierarchical structure in sequences: A linear-time algorithm. *Journal of Artificial Intelligence Research* **1997**, *7*, 67–82. https://doi.org/10.1613/jair.374.
38. Mitarai, S.; Hirao, M.; Matsumoto, T.; Shinohara, A.; Takeda, M.; Arikawa, S. Compressed pattern matching for SEQUITUR. In Proceedings of the Proceedings DCC 2001. Data Compression Conference. IEEE, 2001, pp. 469–478. https://doi.org/10.1109/DCC.2001.917178.
39. A.A. Sekh.; D.P. Dogra.; S. Kar.; P.P. Roy. Video trajectory analysis using unsupervised clustering and multi-criteria ranking. *Soft Computing* **2020**, *24*, 16643–16654. https://doi.org/10.1007/s00500-020-04967-9.
40. Rao, S.; Sastry, P.S. Abnormal activity detection in video sequences using learnt probability densities. In Proceedings of the TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, 2003, Vol. 1, pp. 369–372 Vol.1. https://doi.org/10.1109/TENCON.2003.1273347.
41. Guo, D.; Wu, E.Q.; Wu, Y.; Zhang, J.; Law, R.; Lin, Y. FlightBERT: binary encoding representation for flight trajectory prediction. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *24*, 1828–1842. https://doi.org/10.1109/TITS.2022.3219923.
42. Zhou, B.; Wang, Y.; Yu, G.; Wu, X. A lane-change trajectory model from drivers' vision view. *Transportation Research Part C: Emerging Technologies* **2017**, *85*, 609–627. https://doi.org/10.1016/j.trc.2017.10.013.
43. Gamba, P.; Mecocci, A. Perceptual grouping for symbol chain tracking in digitized topographic maps. *Pattern Recognition Letters* **1999**, *20*, 355–365. https://doi.org/10.1016/S0167-8655(99)00003-3.
44. García-Huerta, J.M.; Jiménez-Hernández, H.; Herrera-Navarro, A.M.; Hernández-Díaz, T.; Terol-Villalobos, I. Modelling dynamics with context-free grammars. *Video Surveill. Transp. Imaging Appl. 2014* **2014**, *9026*, 6. https://doi.org/10.1117/12.2040973.
45. Rosani, A.; Conci, N.; Natale, F.G.D. Human behavior recognition using a context-free grammar. *Journal of Electronic Imaging* **2014**, *23*, 033016. https://doi.org/10.1117/1.JEI.23.3.033016.
46. Hugo, J.H.; Jose-Joel, G.B.; Teresa, G.R. Detecting abnormal vehicular dynamics at intersections based on an unsupervised learning approach and a stochastic model. *Sensors* **2010**, *10*, 7576–7601. https://doi.org/10.3390/s100807576.
47. Xue, Y.; Zhao, J.; Zhang, M. A watershed-segmentation-based improved algorithm for extracting cultivated land boundaries. *Remote Sensing* **2021**, *13*, 939. https://doi.org/10.3390/rs13050939.
48. Zhang, H.; Tang, Z.; Xie, Y.; Gao, X.; Chen, Q. A watershed segmentation algorithm based on an optimal marker for bubble size measurement. *Measurement* **2019**, *138*, 182–193. https://doi.org/10.1016/j.measurement.2019.02.005.
49. Venu, D.; Kumar, A.A. Comparison of Traditional Method with watershed threshold segmentation Technique. *The International journal of analytical and experimental modal analysis* **2021**, *13*, 181–187. https://doi.org/18.0002.IJAEMA.2021.V13I1.200001.015685901874.
50. de la Higuera, C. Grammatical inference: learning automata and grammars. In Proceedings of the Cambridge University Press, 2010, pp. 1–502. https://doi.org/10.1017/CBO9781139194655.
51. Nevill-Manning, C.G.; Witten, I.H. Compression and Explanation using Hierarchical Grammars. *The Computer Journal* **1997**, *40*, 103–116. https://doi.org/10.1093/comjnl/40.2_and_3.103.
52. Åke Björck. *Least Square Method*; Vol. 1, *Handbook of Numerical Analysis*, Elsevier, 1990; pp. 465–652. https://doi.org/10.1016/S1570-8659(05)80036-5.
53. Gomila, R. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General* **2021**, *150*, 700. https://doi.org/10.31234/osf.io/4gmbv.
54. Björck, Å. *Numerical methods for least squares problems*; SIAM, 2024. https://doi.org/10.1137/1.9781611971484.