

Language Identification of Controlled Systems: Modeling, Control, and Anomaly Detection

J. F. Martins, J. A. Dente, A. J. Pires, and R. Vilela Mendes

Abstract—Formal language techniques have been used in the past to study autonomous dynamical systems. However, for controlled systems, new features are needed to distinguish between information generated by the system and input control. We show how the modeling framework for controlled dynamical systems leads naturally to a formulation in terms of context-dependent grammars. A learning algorithm is proposed for on-line generation of the grammar productions, this formulation being then used for modeling, control, and anomaly detection. Practical applications are described for electromechanical drives. Grammatical interpolation techniques yield accurate results, and the pattern detection capabilities of the language-based formulation makes it a promising technique for the early detection of anomalies or faulty behavior.

Index Terms—Control systems, fault diagnosis, formal languages, grammatical inference mechanisms, grammatical interpolation, induction motor drives.

I. INTRODUCTION

Formal language theory has been used in the past to study autonomous systems, the grammatical complexity of the sequences generated according to some coding being used to characterize the complexity of the dynamical system. References [1]–[10] are a representative sample of some of this work. However, most systems used in technological applications are not autonomous, but controlled dynamical systems. Therefore, one has to distinguish, from the start, between the information that is generated by the dynamics of the system and the one that depends on the free will of the controlling operator. This has some implications on the nature of the formal entities that are used to code the system variables and on the nature of the languages that are generated. Another important requirement for practical applications is the possibility to extract the grammars by on-line learning from the data generated by the system, independently of whether or not an analytic model is known.

In Section II, after a short introduction to some formal language concepts, we show how the modeling framework for controlled dynamical systems leads, almost uniquely, to a context-dependent grammatical formulation. A learning algorithm is then proposed for on-line generation of the productions of the grammar. The following sections illustrate how this formulation may be used for modeling, control, and anomaly detection.

Most of our applications concern electromechanical drives. Using grammatical interpolation, as described in Section III, one may obtain quite accurate results, even with small grammar. However the main capabilities of a language-based formulation lies in the pattern detection domain. Therefore, early detection of anomalies or faulty behavior is probably the most promising domain of application for this technique.

Manuscript received April 13, 2000; revised May 15, 2001. This paper was recommended by Associate Editor A. Zomaya.

J. F. Martins and A. J. Pires are with the Laboratório de Mecatrónica, Instituto Superior Técnico, Lisboa, Portugal and with the Escola Superior de Tecnologia de Setúbal, Setúbal, Portugal.

J. A. Dente and R. Vilela Mendes are with the Laboratório de Mecatrónica, Instituto Superior Técnico, Lisboa, Portugal.

Publisher Item Identifier S 1094-6977(01)06491-4.

II. GRAMMARS FOR CONTROLLED DYNAMICAL SYSTEMS

A. Formal Language Concepts

Many fundamental contributions to the formal language field originate from the work of [11]–[14], whose theory of formal grammar had a major influence in the development of the subject. *Grammatical inference*, that is, the development of algorithms which extract grammatical information from examples, is a concept that goes back to Gold's work [15]. Since then, a great deal of work has been done, which can be found in several surveys [16]–[18], [25].

To apply grammatical inference procedures, a dynamical system must be considered as an entity (linguistic source) capable of generating a specific language. The *grammar* G of the language is the set of rules that specifies all the *words* in the language and their relationships. Once the grammar is found, the grammar itself is a *model* for the source. To define a grammar G , one specifies a *terminal alphabet*, a *nonterminal alphabet*, a *start symbol*, and a *set of productions*

$$G = \{\Sigma_T, \Sigma_N, S, P\}. \quad (1)$$

- 1) *Terminal Alphabet* (Σ_T): Set of symbols that make up the words, a word being a string of symbols.
- 2) *Nonterminal Alphabet* (Σ_N): Set of auxiliary symbols that are used to generate the words by the production rules.
- 3) *Start Symbol* (S): A special nonterminal symbol used to start the generation of words.
- 4) *Productions* (P): Set of substitution rules (denoted $a \rightarrow b$) used to generate the allowed words in the language.
- 5) *Example*: The grammar.

$$\{\Sigma_T = (a, b), \Sigma_N = (A), P = (S \rightarrow bA, A \rightarrow aA, A \rightarrow a)\} \quad (2)$$

generates the language with words consisting of a b symbol followed by any number of a symbols.

The set of productions encodes the dynamics of the system that generates the language. Any word that can be derived from the start symbol by a sequence of productions of the grammar is said to be in the language generated by the dynamical system.

Grammatical inference is an algorithm by which a grammar is inferred from a set of sample words produced by the dynamical system considered as the linguistic source. Notice that there is not a unique relation between a language and the grammar because distinct grammars may generate the same language.

Grammatical inference, in general, is the identification of a grammar from a set of positive and negative examples. One may also consider a scheme where a “teacher” answers questions concerning the language to be inferred. However, in dynamical system identification by grammatical inference, we are in the restricted setting of language identification from positive examples alone. Therefore, a finite sample cannot characterize an infinite language. It may at most specify the class of languages, which possesses that finite sample as an allowed one.

A practical problem, when learning from a finite set of examples, is the fact that some relevant productions may never be inferred. If a metric is defined in the space of words, this problem may be solved to some extent by *grammatical interpolation techniques*. Another practical problem arises when the system under study is perturbed by noise. Then, productions may be inferred which are not characteristic of the system but are a result of the noise perturbation. The solution used in

our applications is to keep as valid only those productions that appear many times in the learning process. The threshold for acceptance of a production depends of course on the size of the training set and the estimated noise intensity.

We work here in the *phrase-structure grammar* setting, as defined by Chomsky. Notice, however, that more general systems like *programmed grammars*, *cooperating distributed grammars*, or *contextual grammars* [14] may also find useful applications in the characterization of dynamical systems. In particular, programmed languages and the interpretation of controls as labels could be a viable alternative to our identification of the control variables with nonterminal symbols.

B. Grammatical Inference in Controlled Dynamical Systems

A model for a controlled dynamical system has the general form

$$\begin{aligned}\dot{x} &= f(x, U) \\ y &= h(x)\end{aligned}\quad (3)$$

or, considering the time divided into discrete steps

$$\begin{aligned}x_{k+1} &= f(x_k, U_k) \\ y_k &= h(x_k)\end{aligned}\quad (4)$$

where

- x state variable;
- y output (or observed) variable;
- U input (or control) variable.

Equations (3) and (4) also establish a functional relationship between the output variables at different times

$$y_{k+1} = g(y_k, U_k). \quad (5)$$

However, in most systems used in technology, not all state variables are observable. Therefore, (5) does not provide a complete specification of the system. In general, specification of the dynamics in terms of the output variables requires a set of functional relationships involving many time steps in the past, namely

$$\begin{aligned}y_{k+1} &= g_0(U_k) \\ y_{k+1} &= g_1(y_k, U_k) \\ y_{k+1} &= g_2(y_k, y_{k-1}, U_k) \\ y_{k+1} &= g_3(y_k, y_{k-1}, y_{k-2}, U_k) \\ &\vdots\end{aligned}\quad (6)$$

It is this structure, which is required by dynamical considerations on actual controlled systems, that leads in a natural way to our proposal of p -type productions.

To develop a grammatical description and a grammatical inference algorithm for controlled dynamical systems three steps are required. First, the quantification of the variables, then the specification of the nature of the productions and finally a learning algorithm to extract the productions from the experimental data.

Quantification: Quantification refers to the creation of alphabets for the output variable y and the control variable U . In our approach, we associate the terminal alphabet Σ_T to the output variable y and the nonterminal alphabet Σ_N to the control variable U . Let n be the number of terminal symbols and m the number of nonterminal symbols. A quantification of the variables is made, in a discrete way, dividing the variables range in equal intervals and associating each interval to a symbol in the alphabet

$$\begin{aligned}y &\leftrightarrow y_i \in \Sigma_T \quad i = 1 \cdots n \\ U &\leftrightarrow U_j \in \Sigma_N \quad j = 1 \cdots m.\end{aligned}\quad (7)$$

Productions: P -type productions are defined to be substitution rules of the form

$$y_1 \cdots y_p U_k \rightarrow y_1 \cdots y_p y_{p+1} \delta \quad (8)$$

where

- $y_1 \cdots y_p$ sequence of terminal symbols;
- U_k nonterminal symbol;
- δ special nonterminal.

δ is used to allow the conclusion, or not, of a generated word, by the use of the following special set of productions

$$\begin{aligned}\delta &\rightarrow U_j, j = 1 \cdots m \\ \delta &\rightarrow \lambda\end{aligned}\quad (9)$$

where λ denotes the empty symbol.

A p -type production codes the evolution of the output variable, depending on its p past values and on the value of the control variable U . There is, therefore, a functional relationship between the dynamics of the system and the p -type productions.

Learning Algorithm: To obtain a sample of the language, a sequence of control signals is applied to the system in such a way that the output variable y takes values in a sufficiently wide region. The signal evolution is then quantified as described above, and the following learning algorithm is used.

- 1) A 0-type production is assumed for every newly occurring control symbol.
- 2) A new $(n + 1)$ -type production is generated each time the data conflicts with the previously established n -type productions. The conflicting n -type productions are also promoted to $(n + 1)$ -type productions or are deleted if there is not sufficient information in the past to do so.

To be able to revise the conflicting n -type productions, promoting them to $(n + 1)$ -type productions, we need to keep in a memory window the record of a certain number of past steps in the system evolution. The length of the memory window need not be larger than p_{\max} (p_{\max} being the highest order allowed for the productions). In general, there is not great harm in having zero-length for the memory window and just deleting all the conflicting lower order productions. This is because all the most relevant productions will reappear some time again in the future.

As an example, consider the following symbol sequence obtained from a data sample

$$\begin{aligned}U \text{ variable} &: A B A A B A \\ y \text{ variable} &: e d c b d e.\end{aligned}\quad (10)$$

At time zero, the algorithm analyzes the leading symbols of both the control and the output variable. Since no other information is yet available, a 0-type production $A \rightarrow e\delta$ is assumed. After analyzing the second symbol, the algorithm establishes another 0-type production $B \rightarrow d\delta$. The third symbol would yield a 0-type production $A \rightarrow c\delta$. However, this production contradicts the previously established $A \rightarrow e\delta$ because the grammar is not stochastic. Therefore, in this case, a 1-type production $dA \rightarrow dc\delta$ is obtained. The 0-type production $A \rightarrow e\delta$ is deleted because no information is available in the past on the first symbol (e). At the next step, we obtain the production $cA \rightarrow cb\delta$. Another conflict arises when we reach the last symbol in (10). Then a 2-type production $bdA \rightarrow bde\delta$ is obtained and $dA \rightarrow dc\delta$ is revised to $edA \rightarrow edc\delta$. Finally, one is left with

$$\begin{aligned}B &\rightarrow d\delta \\ cA &\rightarrow cb\delta \\ bdA &\rightarrow bde\delta\end{aligned}$$

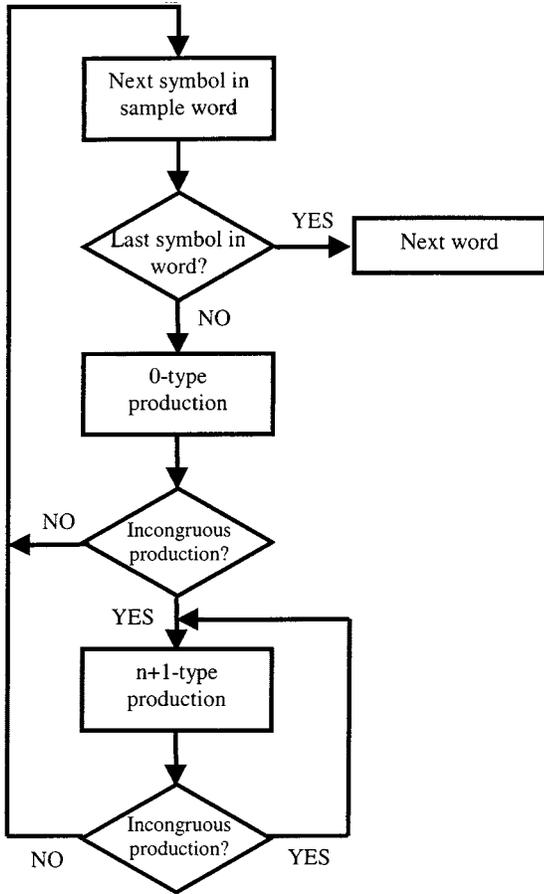


Fig. 1. Flowchart of the grammatical inference learning algorithm.

$$edA \rightarrow edc\delta. \tag{11}$$

The learning algorithm is described by the flow chart in Fig. 1. Unlike, for example, the rules for fuzzy logic inference, the structure of the formal language productions, obtained by our algorithm, is not established in advance. A mixture of different types of productions may be obtained in the grammar. This feature provides flexible adaptation in dynamical systems displaying distinct behavior in different regions of the working space.

In a controlled system, the two types of variables, output variables and control variables, have a distinct nature. This is represented in our formalism by the assignment of terminal symbols to the dynamical variables and nonterminal symbols to the control. In turn, the productions represent the action of the control in the context of the past and present dynamics of the system. The languages generated in this way fall into the category of context-sensitive languages [12].

In practical applications, the grammatical inference algorithm is completed with *grammatical interpolation* and a *noise rejection* mechanism. Grammatical interpolation is discussed in Section III, and noise rejection is simply implemented by keeping only those productions that appear a sufficient number of times in the data. This simple procedure ensures the stability and robustness of the learning algorithm. Also, for practical purposes, an upper bound is put on the order of the productions that the algorithm generates.

As a simple example of how this methodology is used to model the dynamics, we consider the following piecewise linear system:

$$y_{k+1} = y_k - y_{k+1} + 2U_k \quad \text{with} \quad \begin{cases} y_k \in \{1, 2, 3, 4\} \\ U_k \in \{-1, 0, 1\} \end{cases}. \tag{12}$$

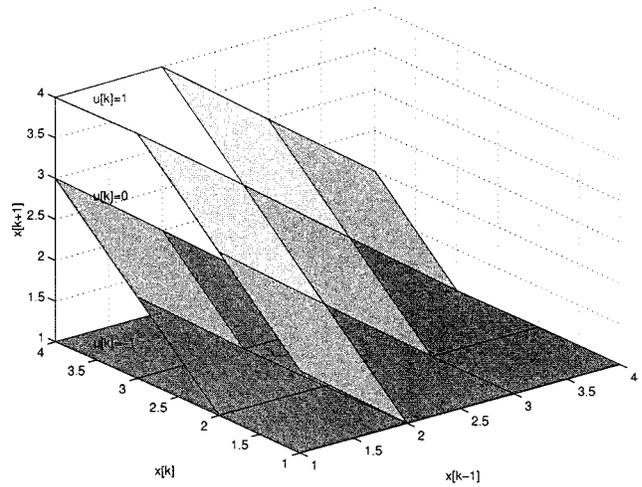


Fig. 2. Working domain of a piecewise linear dynamical system.

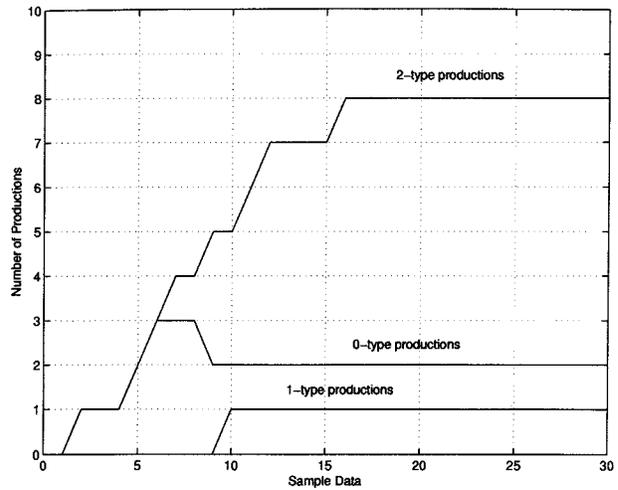


Fig. 3. Evolution of the learned grammar productions for a piecewise linear dynamical system.

The working domain of this dynamical system is spanned by three distinct planes, depending on the values of the U_k variable (Fig. 2). The assumed codification is

$y_k \leftrightarrow \Sigma_T$	$U_k \leftrightarrow \Sigma_N$
$1 \leftrightarrow a$	$-1 \leftrightarrow A$
$2 \leftrightarrow b$	$0 \leftrightarrow B$
$3 \leftrightarrow c$	$1 \leftrightarrow C$
$4 \leftrightarrow d$	

The language generated by this dynamical system is represented by a grammar G_1 , whose rewriting system has one 0-type production, two 1-type productions, and 24 2-type productions.

This system was used to test our on-line learning algorithm with a random input control signal U_k . Fig. 3 shows the evolution of the productions, extracted from randomly generated data, for the first 30 data points. One sees, for example, at the data point 19, the cancellation of a 0-type production forced by a new 1-type production.

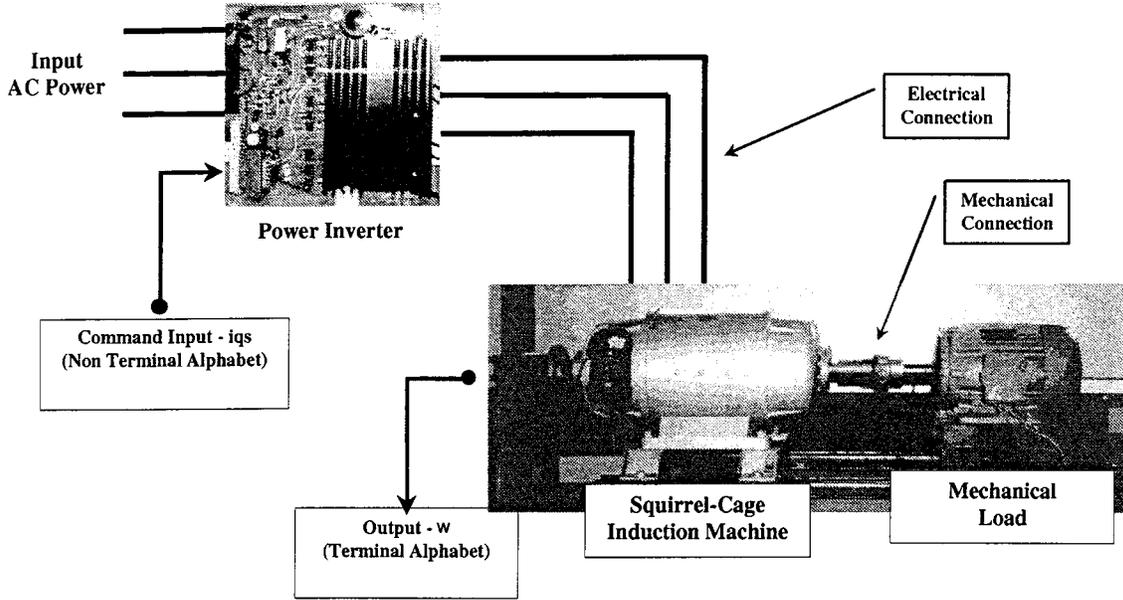


Fig. 4. Electrical drive system.

III. MODELING AN ELECTRICAL DRIVE

The linguistic coding described before has been used, in our laboratory, to model an experimental drive system, consisting of an electronically fed induction machine.

The experimental system is depicted in Fig. 4. It is composed of an induction motor driven by a power inverter. Drive systems of this type are modeled using the electromechanical power conversion theory [19]. A 12-equation model represents the electrical drive system. The large number of variables hinders any attempt to perform on-line learning without huge computational costs. Therefore, one needs to simplify the electrical drive model.

Considering an input variable internal control loop [20], stator currents may be assumed to be controlled. The model is then reduced from a twelfth-order to a third-order system [21], which in the rotor flux reference frame, is

$$\begin{aligned}\dot{\psi}_{dr} &= -\frac{1}{\tau_r} \psi_{dr} + (\omega_r - \omega) \psi_{qr} + \frac{M}{\tau_r} i_{ds} \\ \dot{\psi}_{qr} &= -(\omega_r - \omega) \psi_{dr} - \frac{1}{\tau_r} \psi_{qr} + \frac{M}{\tau_r} i_{qs} \\ \dot{\omega} &= \frac{M}{JL_r} (i_{qs} \psi_{dr} - i_{ds} \psi_{qr}) - \frac{B}{J} \omega - T_{ext}.\end{aligned}\quad (13)$$

The rotor fluxes (ψ_{dr} , ψ_{qr}) and the rotor speed (ω) are the state variables and the stator current components (i_{ds} , i_{qs}) the control variables. In addition

- 1) τ_r rotor time constant;
- 2) M mutual (stator-rotor) induction coefficient;
- 3) J inertia coefficient;
- 4) L_r rotor self-inductance coefficient;
- 5) B friction coefficient;
- 6) T_{ext} load torque.

With the reference frame rotor flux given by

$$\omega_r = \omega + \frac{M}{\tau_r} \frac{i_{qs}}{\psi_r}\quad (14)$$

one obtains from (13)

$$\dot{\psi}_{qr} = -\frac{1}{\tau_r} \psi_{qr} + \frac{M}{\tau_r} i_{ds}$$

$$\dot{\omega} = \frac{M}{JL_r} i_{qs} \psi_r - \frac{B}{J} \omega - T_{ext}.$$

To obtain relevant training sets, two types of information qualitative and quantitative should be acquired. Qualitative information is obtained from the mathematical models. This leads to the choice of the relevant variables that describe the system behavior. This choice is very important to insure the existence of a functional relationship representative of the drive behavior [22]. Otherwise, both learning and recognition would be impossible.

To obtain quantitative information experimental data is acquired. To obtain the data an excitation signal must be chosen. A possible approach would be to use a pseudo-random binary signal [23]. However, this signal is not the best choice for drive systems because it is filtered by mechanical time constants. It is better to use sinusoidal signals of different amplitudes and frequencies. In this way, frequencies and amplitudes may be adjusted within the limits of the drive response, thus avoiding the filtering problem and collecting data that adequately spans the operating domain.

A grammar of the drive language is inferred from the (control) input/output (I/O) experimental information. The i_{ds} current is kept constant, a nonterminal alphabet is established from the quantification of the i_{qs} current and a terminal alphabet established from the output speed signal (ω). A representative training set is obtained with a sinusoidal reference signal containing a combination of different amplitudes and frequencies. Fig. 5 shows the evolution, in the I/O space, of the training and test data sets, with the test data being displayed in bold.

The acquisition of the training set is an essential step in obtaining a good knowledge of the system to be modeled. The training set must cover a representative part of the entire working domain.

A. Recognition Results

To test the validity of the recognition process, two alphabets were considered. A quantification is established for both input and output variables, yielding a 60-symbol alphabet.

The inferred grammar contains 183 productions distributed in the following way: five 0-type productions, 156 1-type productions, and 22 2-type productions. To avoid the influence of noise on the inferred productions, a production is considered valid only if it appears more

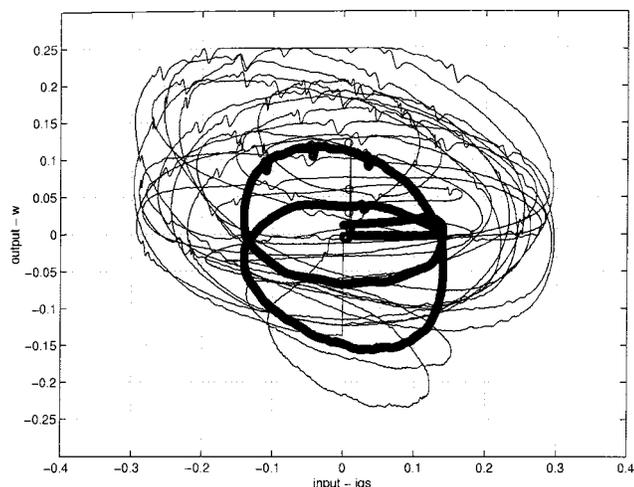


Fig. 5. Training and test data set for the electrical drive.

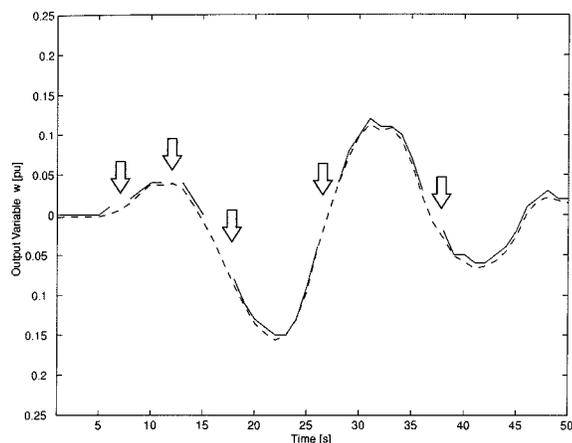


Fig. 6. Formal language recognition (60-letter alphabet).

than a certain number of times, this parameter depending on the size of the training set. This procedure works as a filter on noise perturbations.

Applying this grammar to the test set, we obtain the recognition results shown in Fig. 6. The actual electromechanical drive speed is shown as a dotted line and the grammar response as a continuous one.

Apart from the quantification error, the recognition results may be considered satisfactory because the evolution of the drive speed and the grammar results are similar. However, we must point out that there are some situations, indicated by the white arrows in Fig. 6, where the grammar does not provide any answer. This happens when no production was inferred representing that particular I/O relationship. As explained later, a generalization method is used to establish nonlearned productions.

For the second alphabet, a quantification interval ten times higher is assumed for both input and output variables, yielding now a six-symbol alphabet, instead of the previous 60 symbols. The inferred grammar contains 36 productions, distributed in the following way: zero 0-type productions, four 1-type productions, and 32 2-type productions. Applying this grammar to the test set the recognition results are those shown in Fig. 7. As before, the electromechanical drive speed is shown as a dotted line and the grammar response as a continuous one. The white arrows denote the absence of applicable productions.

As expected, the quantitative modeling performance deteriorates, there being a larger difference between the measured drive speed and the grammatical reconstruction. However, if the important feature is the

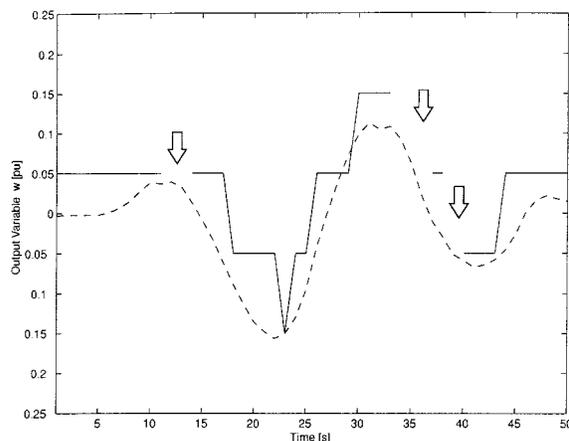


Fig. 7. Formal language recognition (six-letter alphabet).

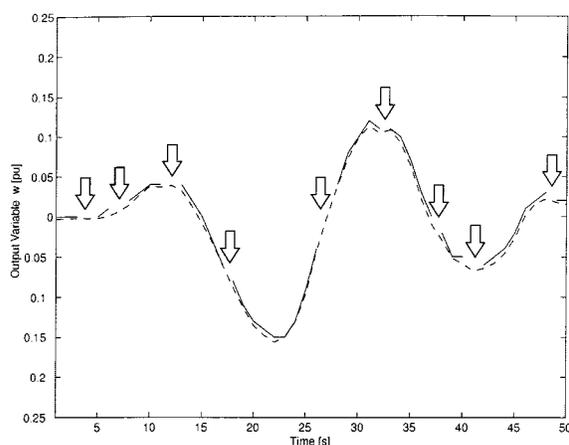


Fig. 8. Formal language recognition with a poor training set.

qualitative recognition of the drive speed, the results may still be considered as satisfactory. For a qualitative recognition process a limited alphabet is not a serious drawback. However, as in the larger grammar, the problem of nonexistent productions is present here.

The discrete quantification performed within the grammatical inference algorithm implies a discrete response when recognition of the drive is performed. In order to improve the quantitative recognition capabilities of the algorithm, two solutions may be used. The first is to increase the number of symbols in the alphabet to make a more accurate quantification. Alternatively, one may use numerical interpolation of the productions. The first solution increases the number of productions thus slowing the recognition process. The second solution does not increase the number of productions.

To illustrate the operation of the interpolation process consider again the 60-symbol alphabet and another training set which is half the size of the one used for the results in Fig. 6. After inferring the correspondent grammar, the recognition results are shown in Fig. 8. It is seen that the recognition process has considerable faults because, for the reduced training set, the number of productions that is obtained is considerably smaller. A smaller training set led to a faulty grammar, which fails to recognize some words. However, because the symbols in our alphabets are in correspondence with numerical values of the variables domain, a metric may be defined in the space of words. New productions may then be obtained, from those inferred from the smaller sample, by interpolation.

Grammatical interpolation establishes new productions by a structural matching procedure. The main idea of structural matching is based

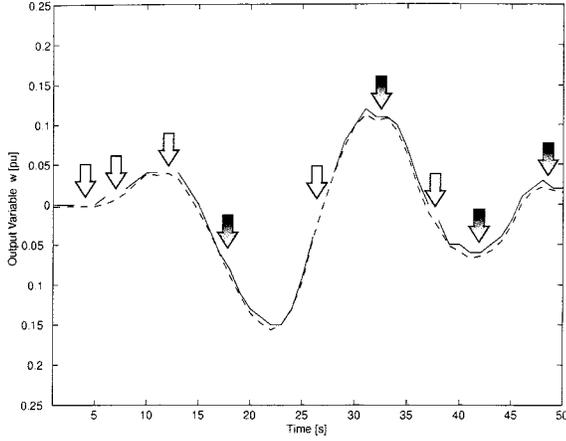


Fig. 9. Formal language recognition with grammatical interpolation.

on a measure of similarity between the unknown input pattern and the available data structures. The measure of similarity is the distance between the nonexistent production and the nearest productions. Several methods for structural word matching have been reported in the literature [24]. The basic algorithm states that the distance between two words is related to the sequence of edit operations (substitution, insertion, and deletion) required to transform one word into another. For any sequence of edit operations, a cost function $c(s)$ is considered

$$c(s) = \sum_{i=1}^n c(e_i)$$

where $c(s)$ denotes the cost of a particular sequence s , and $c(e_i)$ the cost of a particular edit operation. The distance between two words x and y is defined as the minimum cost of transforming a word into another

$$d(x, y) = \min\{c(s) | s \text{ being a sequence of operations transforming } x \text{ into } y\}.$$

When a word cannot be recognized because there is no production generating some symbol in that word, a grammatical interpolation formula is used to obtain the last terminal symbol y_{p+1} in the right-hand side of the new production

$$y_{p+1} = \text{quantification} \left(\left(\sum_{i=1}^n d_i \right) \sum_{i=1}^n \frac{1}{d_i} y_{p+1}|_{\text{prod } i} \right).$$

This interpolation formula establishes a weighed average of the available productions of the required order, on an average distance between similar productions. The distance d_i is the distance between the words $y_1 \cdots y_p U_k$ in the left-hand side of the existent productions and the word of the new production.

Fig. 9 shows the recognition results for the same grammar as considered in Fig. 8 when the grammatical interpolation procedure is applied. The effect of the nonexistent productions is reduced, and the number of symbols that are not recognized are much smaller. Grey arrows denote the interpolation of nonexistent productions, and white arrows productions that could not be obtained by interpolation.

By *generalization* we mean the ability of an algorithm to perform even under working conditions distinct of those observed during the learning phase. The interpolation process as described above provides generalization abilities to the algorithm. However, this type of generalization is a local effect in the working space. For example, consider the training set shown in Fig. 10.

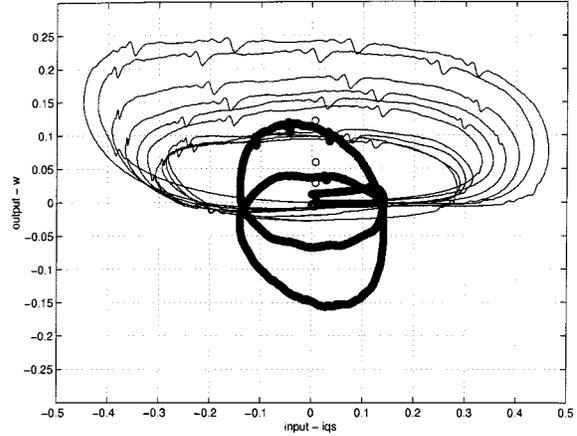


Fig. 10. Restricted training and test data set for the electrical drive.

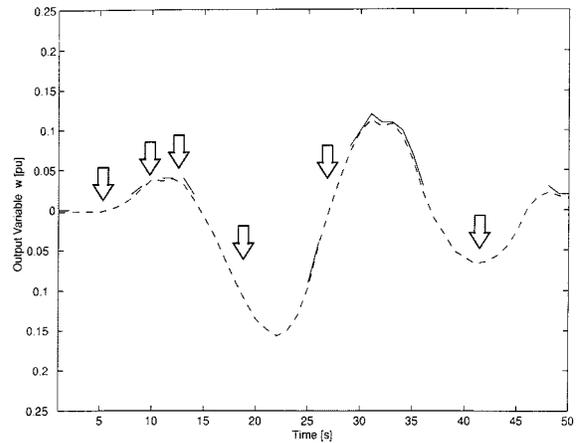


Fig. 11. Generalization using grammatical interpolation.

In this case, the training set contains only positive speed values. Fig. 11 shows the modeling results that are obtained using this sample data. Recognition fails for some of the domain areas not covered by the training data set. Namely, symbols that code for negative drive speed are not recognized. Generalization has only a local effect. Therefore, experimental training data must always cover a significant part of the working domain.

IV. ANOMALY DETECTION

Any grammar G codes for the class of patterns that belong to the language $L(G)$ the grammar generates. In this way, one may use the grammar to recognize well-formed strings and reject anomalous ones. Once one has learned the grammar of the strings generated by the system under normal conditions, if at a later time there is some fault in the dynamic system, strings that will be generated are not compatible with the learned grammar. It is also reasonable to assume that this might occur already at an early stage of the anomaly. The language algorithm might therefore be valuable as a tool for early fault detection.

Detection of grammar anomalies is based on the same distance and similarity measures as considered before for grammar interpolation. The anomaly detection algorithm (Fig. 12) computes the distance between the string generated by the dynamical system and the one generated by the learned grammar. If this distance exceeds some threshold, a fault is reported.

As a first example, let us consider the dynamical system represented by (12), with the same codification and inferred grammar as before.

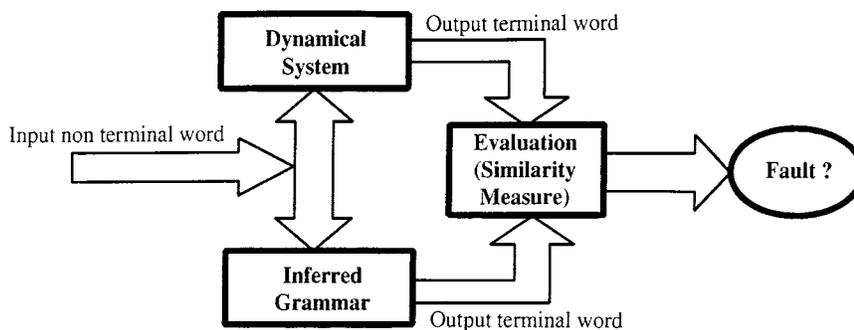


Fig. 12. Fault detection.

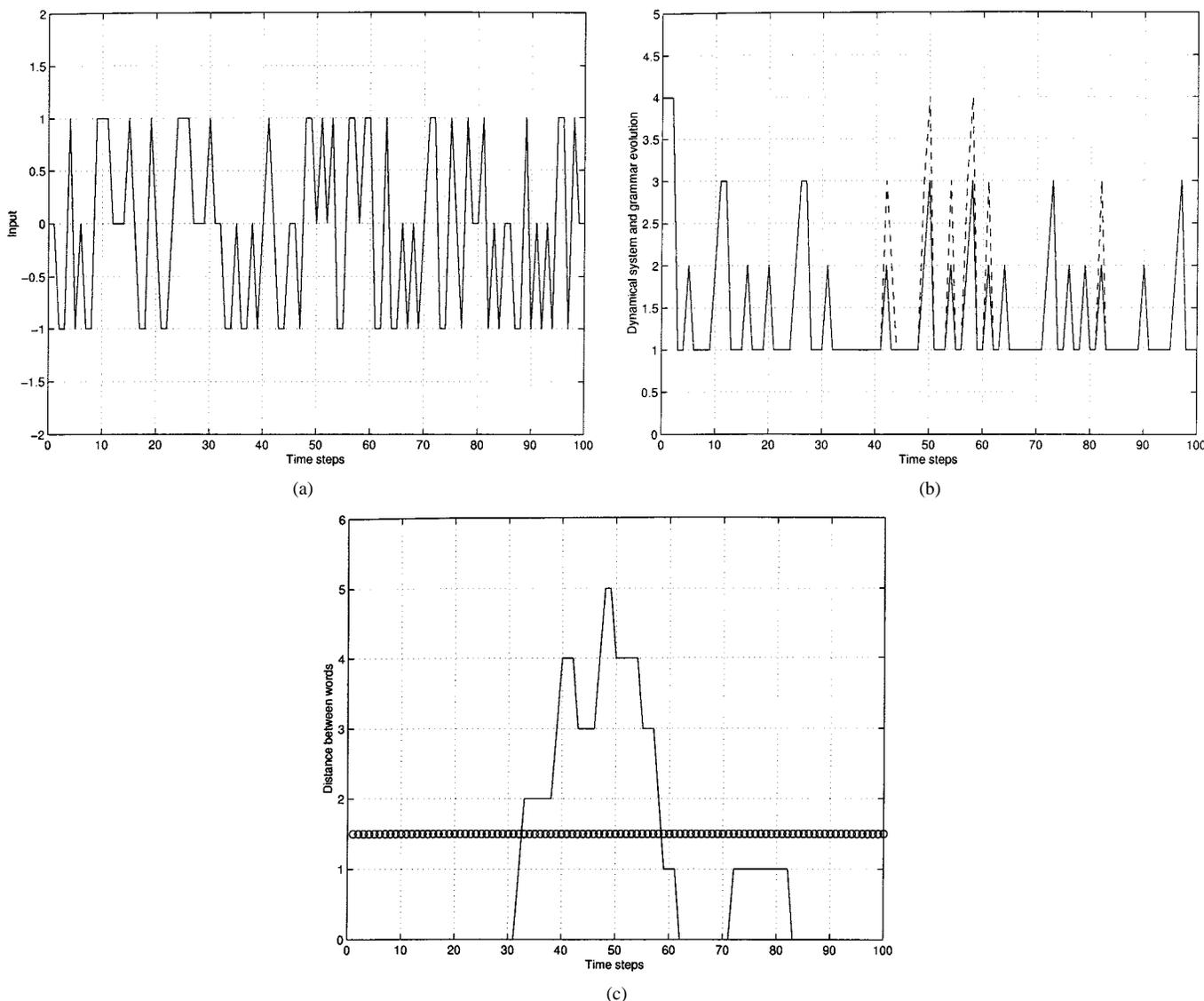


Fig. 13. (a) Control variable. (b) Learned grammar and actual system outputs. (c) Distance between words.

For the anomaly detection test, we consider a random input during 100 time steps. A sustained anomaly was simulated between time steps 30 to 60 and a spurious one between time steps 80 to 85. The anomaly simulation is simply the replacement of the parameter 2 by 3 in (12). The dynamical system changes and, therefore, some of the strings that are generated do not match the learned grammar. Whenever the distance between the system-generated strings and those of the grammar exceeds a threshold, an anomaly is reported.

Fig. 13(a) displays the anomaly detection results. In response to the control variable both the learned grammar and the system produce terminal words. These are shown in Fig. 13(b), with the output of the grammar as a continuous line and the output of the dynamical system as a dashed line. They differ whenever there is an anomaly. Fig. 13(c) displays the distance between words produced by the learned grammar and those of the dynamical system for a word length of ten symbols. This distance exceeds the threshold only for the sustained anomaly.

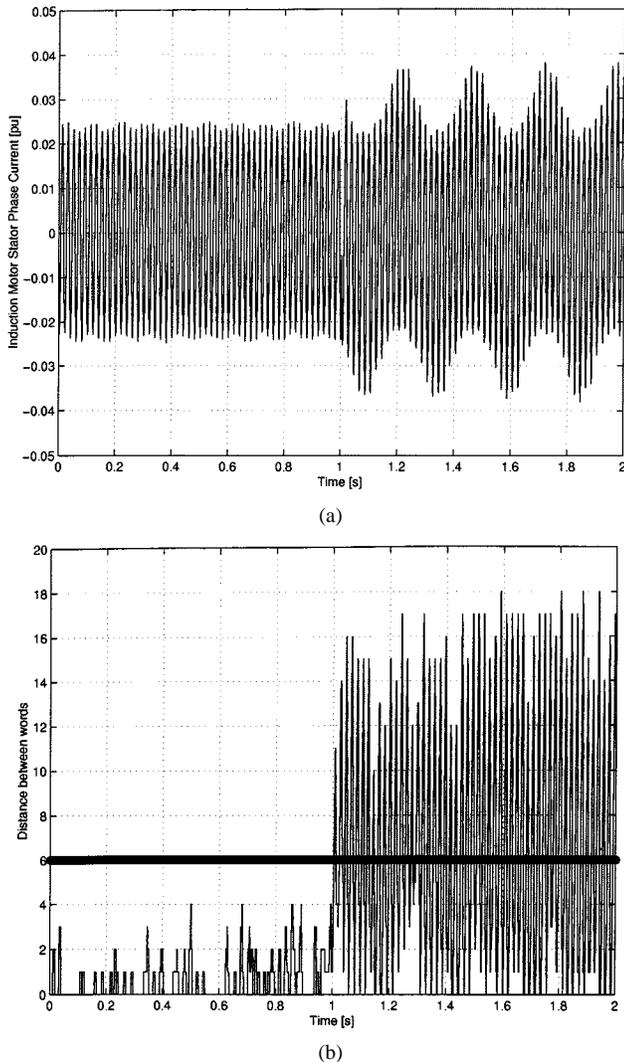


Fig. 14. (a) Induction motor stator phase current. (b) Distance between words.

As a second example, we consider anomaly detection in an induction motor. The anomaly is a rotor broken bar, and the variable coded as a terminal alphabet is the stator phase current. The learned grammar, representing the experimental drive system, is generated with the motor operating without anomaly. Fig. 14(a) shows the stator current deviation due to the presence of the anomaly between time 1 and 2 (s). Fig. 14(b) shows the distance between the words generated by the learned grammar and those of the induction motor. Soon after the onset of the anomaly, this distance begins to exceed the threshold. A nonzero threshold must be defined because, as is seen in the figure, some deviations are obtained even without anomalies due to noise or even to small mismatches in the learned grammar.

V. CONTROL

Once the grammar generated by a dynamical system is learned, this grammar, defining the structural features of the words produced by the linguistic source, is a model for the source. So it is possible to use the productions to predict the evolution of the system or to find the control string that leads from one state to another.

The basic idea of the control algorithm is as follows. A string of terminal symbols is given which represents the sequence of states that one wants the system to follow. Knowing the initial state of the system, for each symbol of the desired sequence of target states, one chooses

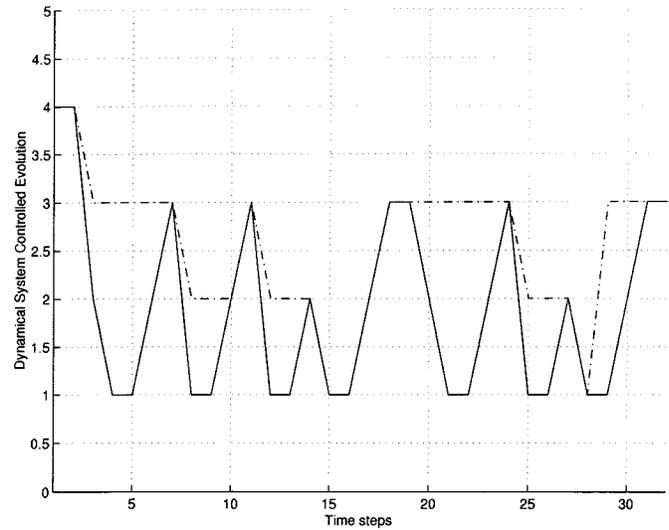


Fig. 15. Controlled evolution of a dynamical system.

a production leading the system from its present state to the desired one. At each step the production that is chosen is the one that leads the system as close as possible to the desired state. The procedure is repeated until that symbol is reached, and then, one moves to the next target symbol.

As an example, consider the dynamical system defined by (12). Suppose that the sequence of target states is “*cbcbaabcccbacc*.” Fig. 15 shows the evolution of the system obtained by the control methodology described above. The target sequence is shown as a dashed line and the system evolution as a continuous line.

Notice that to evolve from one target symbol to another the system may have to pass through other terminal symbols. The control algorithm is robust in the sense that whenever, at some time t , the expected evolution differs from the one predicted by the learned grammar, due to some perturbation or noise, it suffices to apply the same algorithm once more starting from the state at time t .

VI. CONCLUSIONS AND REMARKS

This paper proposes a characterization of controlled system dynamics by formal language techniques. The applications we have been concerned with are mostly electromechanical drives. However, the technique might also be useful for many other dynamical systems or industrial processes.

An important issue is the choice of an appropriate quantification for the feature space. This influences the choice and size of the alphabets and the dimension and complexity of the grammars. Also, good coverage of the working domain, in the learning stage, is essential to insure good generalization properties.

Unlike other approaches, the nature of the rules that define the dynamical system are not set up in advance, and different types of productions are established on-line according to the incoming words from the linguistic source. The method seems to provide good recognition results and a reasonable quantitative accuracy. Fuzzy logic approaches are probably more efficient when quantitative accuracy is very important, and a good knowledge already exists concerning the rules of the system. However, if the dynamics are unknown or if pattern recognition is the important issue, the formal language approach seems quite promising.

Fault detection in control systems, industrial plants, and power networks is a field where this work is being pursued and compared with immunity-based learning systems.

REFERENCES

- [1] S. Wolfram, "Computational theory of cellular automata," *Commun. Math. Phys.*, vol. 96, pp. 15–57, 1984.
- [2] P. Grassberger, "Toward a quantitative theory of self-generated complexity," *Int. J. Theor. Phys.*, vol. 25, pp. 907–938, 1986.
- [3] J. P. Crutchfield and K. Young, "Computation at the onset of chaos," in *Complexity, Entropy and Physics of Information, SFI Studies in the Sciences of Complexity*, W. H. Zurek, Ed. Reading, MA: Addison-Wesley, 1990, vol. 8, pp. 223–267.
- [4] D. Auerbach and I. Procaccia, "Grammatical complexity of strange sets," *Phys. Rev.*, vol. A41, pp. 6602–6614, 1990.
- [5] G. Troll, "Formal languages in dynamical systems," *Acta Univ. Carolina Math. Phys.*, vol. 34, p. 117, 1993.
- [6] H. Xie, "On formal languages in one-dimensional dynamical systems," *Nonlinearity*, vol. 6, pp. 997–1007, 1993.
- [7] Y. Wang and H. Xie, "Grammatical complexity of unimodal maps with eventually periodic kneading sequences," *Nonlinearity*, vol. 7, pp. 1419–1436, 1994.
- [8] P. Collet, A. Galves, and A. Lopes, "Maximum likelihood and minimum entropy identification of grammars," *Random Computat. Dyn.*, vol. 3, p. 241, 1995.
- [9] P. Lakdawala, "The computational complexity of symbolic dynamics at the onset of chaos," *Phys. Rev.*, vol. E53, pp. 4477–4485, 1996.
- [10] C. Moore, "Dynamical recognizers: Real-time language recognition by analog computers," Santa Fe Inst., Santa Fe, NM, working paper 96-05-023.
- [11] N. Chomsky, "On certain formal properties of grammar," *Inform. Contr.*, vol. 2, pp. 137–167, 1959.
- [12] ———, *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
- [13] R. N. Moll, M. A. Arbib, and A. J. Kfoury, *An Introduction to Formal Language Theory*. New York: Springer-Verlag, 1988.
- [14] G. Rozenberg and A. Salomaa, *Handbook of Formal Languages*. New York: Springer-Verlag, 1997.
- [15] E. M. Gold, "Language identification in the limit," *Inform. Contr.*, vol. 10, pp. 447–474, 1967.
- [16] D. Angluin, "Inductive inference of formal languages from positive data," *Inform. Contr.*, vol. 45, pp. 117–135, 1980.
- [17] K. S. Fu and T. L. Booth, "Grammatical inference: Introduction and survey—Part I," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-5, pp. 95–111, Jan. 1975.
- [18] Y. Sakakibara, "Recent advances in grammatical inference," *Theoretical Comput. Sci.*, vol. 185, pp. 15–45, 1997.
- [19] A. E. Fitzgerald, C. Kingsley, and S. D. Umans, *Electric Machinery*, 4th ed. New York: McGraw-Hill, 1985.
- [20] J. F. Martins, A. J. Pires, and J. F. Silva, "A novel and simple current controller for three-phase PWM power inverters," *IEEE Trans. Ind. Electron.*, vol. 45, pp. 802–805, Oct. 1998.
- [21] J. F. Martins, A. J. Pires, and J. A. Dente, "Automatic input/output modeling of a squirrel cage induction motor drive system using a neural network," in *Proc. EPE*, Trondheim, Norway, 1997, pp. 4.632–4.637.
- [22] ———, "A choice of variables in automatic modeling of AC drive systems," in *Proc. PEMC*, Prague, Czech Republic, 1998, pp. 5.7–5.12.
- [23] T.-S. Low, T. H. Lee, and H. K. Lim, "A methodology for neural network training from control of drives with nonlinearities," *IEEE Trans. Ind. Electron.*, vol. 40, pp. 243–249, Apr. 1993.
- [24] H. Bunke, "String matching for structural pattern recognition," in *Syntactic and Structural Pattern Recognition. Theory and Applications, Series in Computer Science*, H. Bunke and A. Sanfelui, Eds. Singapore: World Scientific, 1990.
- [25] K. S. Fu and T. L. Booth, "Grammatical inference: Introduction and survey—Part II," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-5, pp. 409–423, Apr. 1975.

An Efficient Gray Search Algorithm for the Estimation of Motion Vectors

Jau-Ling Chen and Pei-Yin Chen

Abstract—Motion vector estimation plays an important role in motion compensated video coding. An efficient and fast search algorithm is proposed for the estimation of motion vector. With the help of gray prediction, the algorithm can determine the motion vectors of image blocks quickly and correctly. Since the proposed algorithm performs better than other search algorithms, such as three-step search (TSS), cross search (CS), new three-step search (NTSS), four-step search (FSS), block-based gradient descent search (BBGDS), simple and efficient search (SES), prediction search (PS), and gray prediction search (GPS), it is very beneficial in applications where video coding speed is important.

Index Terms—Block motion estimation, gray prediction.

I. INTRODUCTION

Recently, several video compression standards, such as H.263 and MPEG, have been proposed for different applications. The standards use motion estimation to reduce the temporal redundancy and use DCT to reduce the spatial redundancy. A block-matching algorithm (BMA) is often used to find the motion vectors of image blocks for its simplicity and regularity. In a typical BMA, a frame is divided into nonoverlapped square blocks of $M \times M$ pels. Then for a maximum motion displacement of d pels per frame, the block of pels in the current frame (called a reference block) is matched against the corresponding blocks (called the candidate blocks) within the square search area of width $M + 2d$ in the previous frame. A straightforward BMA is the full search (FS) algorithm, which exhaustively searches for the best-matched block within the search area to get the optimal motion vector. Due to the high computation requirement of FS, many efficient and fast search algorithms [1]–[8] have been developed in the past.

The motion vector of a block is highly related to the motion vectors of its adjacent blocks in the same image frame due to the spatial (interblock) correlation. In addition, the motion vectors of the currently coding frame are highly related to the motion vectors of the previously processed image frames because of the temporal (interframe) correlation of the image sequences. Using the interblock and interframe correlations, we present an efficient predictive search algorithm for block motion estimation in this paper.

The gray theory, applicable to the prediction problem of a time-varying nonlinear system, was first proposed by Deng [9]. Since the gray model (GM) can make excellent predictions, it has been successfully and widely used in many fields [9]–[11]. The relations among the spatial adjacent motion vectors and among the temporal adjacent motion vectors are neither deterministic nor totally unknown but is partially known (gray). Since the characteristic is very similar to that of a GM, we adopt GM (1, 1), a kind of single variable and first-order linear dynamic GM to implement the interblock and interframe prediction. The experimental results show that the proposed algorithm performs better than other search algorithms, such as the three-step search (TSS) algorithm [1], the cross search (CS) algorithm [2], the new three-step search (NTSS) algorithm [3], the four-step search (FSS) algorithm [4],

Manuscript received September 15, 2000; revised May 15, 2001. This work was supported in part by the National Science Council, R.O.C., under Grant NSC-89-2213-E-218-036. This paper was recommended by Associate Editor I. Gu.

The authors are with the Department of Electronic Engineering, Southern Taiwan University of Technology, Tainan, Taiwan, R.O.C.

Publisher Item Identifier S 1094-6977(01)06492-6.