Using Grammars for Pattern Recognition in Images: A Systematic Review

RICARDO WANDRÉ DIAS PEDRO, FÁTIMA L. S. NUNES,
and ARIANE MACHADO-LIMA, School of Arts, Sciences and Humanities, University of Sao Paulo, Brazil

Grammars are widely used to describe string languages such as programming and natural languages and, more recently, biosequences. Moreover, since the 1980s grammars have been used in computer vision and related areas. Some factors accountable for this increasing use regard its relatively simple understanding and its ability to represent some semantic pattern models found in images, both spatially and temporally. The objective of this article is to present an overview regarding the use of syntactic pattern recognition methods in image representations in several applications. To achieve this purpose, we used a systematic review process to investigate the main digital libraries in the area and to document the phases of the study in order to allow the auditing and further investigation. The results indicated that in some of the studies retrieved, manually created grammars were used to comply with a particular purpose. Other studies performed a learning process of the grammatical rules. In addition, this article also points out still unexplored research opportunities in the literature.

Categories and Subject Descriptors: F.4.2 [**Mathematical Logic and Formal Language**]: Grammars and Other Rewriting Systems—*Decision problems*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis; I.4.10 [**Image Processing and Computer Vision**]: Image Representation; I.5.1 [**Pattern Recognition**]: Models—*Structural*

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Image grammars, computer vision, image representation, formal languages, syntactic methods, pattern recognition

ACM Reference Format:
Pedro, R. W. D., Nunes, F. L. S., and Machado-Lima, A. 2013. Using grammars for pattern recognition in images: A systematic review. ACM Comput. Surv. 46, 2, Article 26 (November 2013), 34 pages.
DOI: http://dx.doi.org/10.1145/2543581.2543593

1. INTRODUCTION

The theory of formal languages was developed in the mid-1950s and aimed at developing theories related to natural languages. However, it did not take long to realize that this theory was also important for studies on artificial languages, especially for languages derived from computing and information technology [Sipser 2006]. Since then, the study of formal languages has been widely used in the parsing of programming languages, modeling of logic circuits, and biological systems, among others.

Authors' addresses: School of Arts, Sciences and Humanities, University of Sao Paulo, Av. Arlindo Bettio 1000, 03828-000 Sao Paulo, Brazil; emails: {rwandre, fatima.nunes, ariane.machado}@usp.br. This research was supported by the National Scientific and Technological Development Council (Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Process 559931/2010-7, grants #2010/15691-0 and #2011/50761-2, São Paulo Research Foundation (FAPESP) and National Science and Technology Institute - Medicine Assisted by Computer Science (Instituto Nacional de Ciência e Tecnologia - Medicina Assistida por Computação Científica (INCT-MACC).

DOI: http://dx.doi.org/10.1145/2543581.2543593

An image is a representation of a physical object that can be stored, manipulated, and interpreted in accordance with users' needs. The image processing operations are generally divided into three levels, each one with specific functions: (i) low-level processing, responsible for removing undesirable data and highlighting the important ones; (ii) midlevel processing, responsible for identifying significant forms to this process (this process is also called *segmentation*); and (iii) high-level processing, responsible for linking the image with some database knowledge.

Understanding the image content has always been a central problem in the area of computer vision and pattern recognition [Zhu and Mumford 2006], and a possible approach to solve this problem is the use of syntactic methods. Notably, several published works have made use of grammars in the last decade and, consequently, the theory of formal languages to deal with problems of computer vision and pattern recognition. In particular, these studies have focused on the recognition of specific objects in images, texture recognition, object construction, and image segmentation.

Systematic reviews have used a series of designed and well-defined steps according to a previously established and documented protocol. And as they can be audited, more reliable results are produced, hence rendering them independent of the reviewers who initially evaluated them. This article aims to present and discuss the results of a systematic review to identify the state of the art in the area of computer vision and pattern recognition using grammars. In this context, we surveyed papers that focus on learning or extracting the structure of the grammars from images, as well as those that use grammar-based approaches to perform image recognition.

Chanda and Dellaert [2004] presented a review on the use of grammars in computer vision and pattern recognition. In addition to presenting the main methods used in the literature for modeling and pattern recognition, the authors cite some areas that may benefit from the use of grammars combined with computer vision. This work differentiates the use of a systematic process to conduct the review presented and the work period considered, which extends to 2010. Additionally, we indicate some challenges extracted from the works included in the review, which could be future research opportunities.

In addition to this introduction, this article is divided into the following sections: Section 2 describes the methodology used in this systematic review, Section 3 contains a description of the studies reviewed as well as their comprehensive analysis, Section 4 discusses the results, and lastly, Section 5 contrives the final conclusions of this work.

2. METHODOLOGY

The systematic review was conducted in three phases: (i) planning, in which the research guidelines were based on a protocol; (ii) conduction, which consisted of executing the search and selection of works of interest according to the inclusion and exclusion of criteria defined in the protocol; and finally, (iii) the data extraction step, which enabled us to scrutinize the selected studies in order to understand the state of the art in the area under investigation.

In the protocol set for the planning stage, the following research questions were specified: "What methods are used to infer grammars from images?" and "What techniques are used for image recognition for a given grammar?"

To answer the specified questions, we used the selected keywords *image, grammar, grammatical, syntactic, linguistic, inference,* and *estimation*, and the following databases: IEEE Xplore, ACM Digital Library, and Periódicos Capes.[1] All studies found

[1]This is a virtual library that brings and provides to public institutions of research in Brazil the best of international scientific production: http://www.periodicos.capes.gov.br.

concerning the research area in question were considered, regardless of the year of publication.

The systematic review was carried out from May 2011 to August 2012. The search for materials that addressed pattern recognition in images using grammars returned 242 articles, of which 58 were in the IEEE Xplore database, 107 in the ACM Digital Library, and 77 from the Periódicos Capes. From the total articles collected, after implementing the inclusion and exclusion criteria, 31 articles were selected and analyzed in full, 20 from the IEEE Xplore database, two from the ACM Digital Library, and nine from Periódicos Capes. The criteria used to include a paper in this review were (a) papers that submitted methods to infer grammars from images, (b) papers that explained pattern recognition techniques in images for a given grammar, and (c) papers that used pattern recognition techniques in images such as AND-OR graphs or attributed grammars. Papers addressing video or event detection were excluded, since the main focus of this review was on static images.

In order to enrich the results, new research was performed in July 2012 in the IEEE Xplore database with the following keywords: *image, vision, grammar, grammatical, syntactic, linguistic, hierarchical model, compositional model, reconfigurable model, shock grammar, NOT(video)*. This new search recovered 28 papers, five of which had previously been analyzed. From the 23 remaining papers, 19 were added to this systematic review after applying the inclusion and exclusion criteria. Moreover, two papers that had not been found in the initial exploratory analysis were added. Considering the keywords and the sources defined in our protocol, they had not been retrieved during the systematic review, but due to their importance they were included. The first one is Chen et al. [2006], the first paper that explored the usage of AND-OR graphs in computer vision. The second one is Si and Zhu [2011], in which the authors not only used AND-OR graphs and grammars but also used unsupervised learning in their structure.

The techniques used were extracted from each study, the primitives used, the results obtained in the image recognition using grammars, and techniques employed for learning the grammatical rules. An analysis of the advantages and disadvantages of each method was performed at the end of the literature review. Also, the potential gaps in the state of the art that can be explored were also broached.

3. RESULTS

Table I shows a synthesis of the 52 articles analyzed, including the techniques and methods employed in each one. Subsequently, a global analysis of the works is expounded (Section 3.1), taking into account relevant topics to identify the state of the art. Each of the papers was individually analyzed and divided into categories that are analyzed in Sections 3.2 to 3.8. The division into categories considered both the type of grammar used and the purpose for which it was used in the computer vision area. Table I shows the summary of the papers classified according to the sequence in which they appear in Sections 3.2 to 3.8.

3.1. Global Analysis

This article shows that in recent years, especially in the last decade, there has been much research on the use of syntactic methods for pattern recognition in images. Most of the papers submitted used a grammar, manually created, and capable of recognizing a particular class of images.

Figure 1 shows the evolution of this research area regarding the quantity of published articles. According to the study by Zhu et al. [2009], several aspects contribute to the recent developments in this research area, which were not consolidated in the 1970s, for example: (i) a consistent mathematical and statistical framework to integrate

Table I. Summary of All Articles Analyzed

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Stuckelberg and Doermann 1999]	Object recognition	b	Model similar to stochastic grammars; hidden Markov models	A region, a component of the object, a measure, etc.	Document containing musical notes	Not shown	The proposed model had not yet been concluded
[Wang et al. 2005b]	Object and texture recognition	c	Spatial random tree; stochastic hidden tree models; stochastic context-free grammars	Pixels and regions of the image	Images of Brodatz textures and images of buildings	Not shown	100% of correct classification of textures and more than 97% of correct classification for building images
[Wang et al. 2006]	Object recognition	c	Spatial random tree; stochastic hidden tree models; stochastic context-free grammars	Pixels and regions of the image	600 images of fronts of buildings and textures	Some production rules are shown	True classification rate greater than 97%
[Siskind et al. 2007]	Object recognition	c	Spatial random tree; stochastic context-free grammars; center-surround algorithms (similar to inside-outside)	Image feature vectors	100 photographs taken of cars and houses	Not shown	True classification rate greater than 97%
[Chen et al. 2006]	Object recognition	b, c	Attribute graph grammars (AND-OR graphs); top-down / bottom-up algorithm	Segments, junctions, clothing piece and human body	Images of clothes created by an artist	Not shown	Presentation of generated images
[Zhu et al. 2009]	Object recognition	a, c	Probabilistic Grammar Markov Model; AND-OR graphs; oriented triplets	Oriented triplets	Images taken from Caltech4 and Caltech 10[2]	Inferred (AND-OR graphs)	True classification rate greater than 87.6%

[2]http://www.vision.caltech.edu/html-files/archive.html.

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Reddy et al. 2009]	Object recognition	c	Belief propagation method; stochastic context-free grammars; Markov random field; AND-OR graphs; top-down / bottom-up algorithm	Pixel blocks representing face components (eyes, nose, etc.)	Human faces (AR Face database;[3] Yale Face database;[4] ORL;[5] Image Parsing;[6] Georgia[7])	Not shown	True classification rate equal to 95% relating to the eyes' position in an image
[Han and Zhu 2009]	Object recognition	c	Attribute graph grammars (AND-OR graphs); top-down / bottom-up algorithm	3D planar retangles projected on images	40 images with on average 38 rectangles each	Shown	Presentation of generated images
[Han and Zhu 2005]	Object recognition	c	Attribute graph grammars (AND-OR graphs); top-down / bottom-up algorithm	3D planar retangles projected on images	40 images of man-constructed objects	Shown	Presentation of generated images
[Rothrock and Zhu 2011]	Object recognition	b	AND-OR graphs	Parts of human body and clothes	Images of pedestrians	Not shown	Shown some images with the results
[Parag et al. 2012]	Object recognition	b	AND-OR graph	Low-level image features	GUI	Shown an example	Detection rate: 90% for push button, 100% for menu list, 90% for radio button, and 67% for checkbox
[Si and Zhu 2011]	Object recognition	a, c	AND-OR graphs and grammars	Features extracted from small patches	Text and images of egret, deer, and bikes	Inferred (AND-OR graphs and grammars)	ROC curves and tables comparing the results of the framework with latent SVM
[Jin and Geman 2006]	Object recognition	b	Probabilistic context-sensitive grammar	Set of pixels forming a brick	Images License plates	Not shown	Images with the results

[3]http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html.
4http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
5http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.
[6]http://www.parsing.com/.
[7]http://www.anefian.com/research/face-reco.htm.

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Tylecek and Sara 2011]	Object recognition	b	Attributed context-sensitive stochastic grammar	Small rectangles	Images of facades	Shown an example	Images with the results
[Soltanpour and Ebrahim-nezhad 2010]	Object recognition	a, c	Mixture hidden Markov model; adaptive neuro-fuzzy inference system; oriented triplets	Oriented triplets	Images taken from the Caltech 101 database	Inferred (not shown)	True classification rate equal to 80% using ANFIS and 93% with using MHMM
[Sainz and Sanfeliu 1996]	Object recognition	a	Augmented regular expressions; context-sensitive languages; active grammatical inference	Pixels	Traffic signal images	Inferred	Average error around 0.68% for the automatons found
[Wang and Jiang 2009]	Object recognition	b	Stochastic context-sensitive grammars	Geometric shapes	Aerial images	Shown some production rules	Shown some images with the results
[Shet et al. 2009]	Object recognition	b	Predicate logic; models of uncertainty	Line segments, circles, corners, etc.	USC-CAVIAR[8] database and set of own data	Not shown	ROC curves
[Mas et al. 2005]	Object recognition	b	Adjacency grammars	Line segments and arcs	700 instances of graphic symbols	Shown	True classification rate equal to 87.7%
[Mas et al. 2008]	Object recognition	b	Adjacency grammars	Line segments and arcs	Images captured by electronic pens	Some production rules shown	Presentation of generated images
[Trzupek et al. 2011]	Object recognition	b	ETPL(k) grammatical ghaphs	Line segments and points	20 arterial images	Generation of ETPL(k) graphs	True classification rate equal to 85%
[Trzupek et al. 2009]	Object recognition	b	ETPL(k) grammatical graphs	Line segments and points	Arteries	Generation of ETPL(k) graphs	True classification rate equal to 85%
[Ogiela et al. 2008]	Object recognition	b	Context-free grammars	Line segments	Medical images	Shown	True classification rate equal to 90%

[8]http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/.

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Ogiela et al. 2009]	Object recognition	a, b	Graph grammars; context-free attributed grammars	Gravity centers of the image segments and line segments	Images of bones and human organs	Shown	True classification rate equal to 93%
[Gidas and Zelic 1997]	Object recognition	b	Hierarchical syntactic models; context-free grammars	Parts of objects	Image of simulated tools, like hammers, shovels, etc.	Not shown	Images with the results
[Toshev et al. 2010]	Object recognition	b	Generic context-free grammars	Planar patches	Images of buildings[9]	Shown an example	89.3% in detection of planar patches. 76.2% in detection of parent of each planar patch
[Christensen et al. 1996]	Object recognition	b	Regular grammars	Objects, geometric relationships, temporal discontinuity	1 example describing a sequence of images	Shows some examples of production rules	Description of the events sequence recorded at a sequence of images
[Lin and Fu 1986]	Object recognition	c	Three-dimensional plex grammar	Surfaces	Geometric objects (boxes, for instance)	Shown	Presentation of generated images
[Gao et al. 2000]	Object recognition	b	Combination of syntactic and structural methods for representing the Chinese writing	Traces and line segments	7920 Chinese characters handwritten	Not shown	Grade 0.774 [0-1] for the character shown in the article
[Zaboli and Rahmati 2007]	Object recognition	b	Shock graphs and shock grammar	Vertex of different shock groups	Binary images[10]	Shown an example	Shown in images and tables
[Hingway and Bhurchandi 2011]	Object recognition	b	Shock graphs and shock grammar	Vertex of different shock groups	Binary images[11]	Shown an example	Shown in images and tables

[9]http://www.wrightstatearc.comwright.
[10]http://www.lems.brown.edu/ dmc/main.html.
[11]www.lems.brown.edu/ dmc.

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Siddiqi and Kimia 1996]	Object recognition	b	Shock graphs and shock grammar	Vertex of different shock groups	Several types of images	Shown an example	Shown in images
[Siddiqi et al. 1998]	Object recognition	b	Shock graphs and shock grammar	Vertex of different shock groups	Images of objects in general	Shown	Shown in images and tables shown in article
[Ferreira et al. 2007a]	Texture recognition	a	Wavelet transforms; fuzzy grammars	Features vector	30 different types of texture	Inferred (from predefined fuzzy rules)	Grade greater than 85% for the rule describing the texture
[Ferreira et al. 2007b]	Texture recognition	a	Wavelet transforms; fuzzy grammars	Features vector	Cork samples taken from production environment	Inferred (from predefined fuzzy rules)	True classification rate greater than 90% for homogeneity and greater than 94% for cork base detection
[Ferreira et al. 2009]	Texture recognition	a, b	Wavelet transforms; fuzzy grammars	Features vector	100 cork samples taken from the production environment	Inferred (from predefined fuzzy rules)	True classification rate greater than 90% for homogeneity and greater than 94% for cork base detection
[Luo et al. 2009]	Objects construction	b	Hierarchical grammar; data-driven Markov chain Monte Carlo	Geometric elements (triangles, rectangles, etc.)	80 images of objects extracted from the LHI[12] public database	Not shown	Presentation of generated images
[Subramanian et al. 2005]	Object construction	b	2D tabled matrix grammar	Vector of elements	3 examples for using the technique	Shown	Presentation of generated images

[12]http://www.imageparsing.com/FreeDataOutline.html.

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Schlecht et al. 2007]	Object construction	b	L-system; Markov chain Monte Carlo	Geometric figures (cylinders, ellipses, etc.)	184 images of the *Alternaria fungi*	Some production rules shown	Presentation of generated images
[Prusinkiewicz et al. 1988]	Object construction	b	L-system	Graphic symbols	–	Not shown (L-system)	Presentation of generated images
[Sun et al. 2009]	Object construction	a, b	L-system	Pixels	–	Inferred (L-system)	Presentation of generated images
[Hemberg and OReilly 2004]	Object construction	b	Extension L-system	Pixels and segments	–	Some production rules shown	Presentation of generated images
[Qu et al. 2008]	Object construction	b	L-systems	Not mentioned	Images of tree trunks	Inferred	Shown some images with the results
[Wu and Bian 2009]	Image segmentation	b, c	Stochastic context-free grammars; AND-OR graphs; top-down / bottom-up algorithm	Pixels	Internal and external images	Not shown	The method takes 10-20 min; presentation of generated images
[Hamdi et al. 2012]	Image segmentation	b	Regular grammar	Set of pixels in a specific format	Cardiac images	Shown	Accuracy of 93.22% in the estimation of the segmented areas
[Wang et al. 2005a]	Scaling change	b, c	Primal sketch representation; graph grammars; Markov chain Monte Carlo; Quadtree	Line segments, crosses, connections, etc.	Images of cars, constructions, and faces	Not shown	Presentation of generated images
[Kanungo and Mao 2003]	Layout recognition	b	Stochastic regular grammar	Header block, footer block and column block	Documents	Not shown	More robust algorithm as that used in the comparison

(Continued)

Table I. Continued

Articles	Objectives	Inclusion criteria	Techniques	Primitives	Sample	Grammars	Results
[Shilman et al. 2005]	Layout recognition	a, b	Stochastic context-free grammars	Digits, characters, symbols, lines, etc.	Scientific documents	Inferred and manually created	85% accuracy in recognizing the layout of the document and 94% accuracy in the recognition of mathematical formulas
[Kong et al. 2012]	Layout recognition	b	Context-sensitive graph grammars	Text blocks, buttons, images, etc.	Web pages	Not shown	Graphs displayed as images on the article
[Mao et al. 2003]	Layout recognition	b	Attributed context-sensitive stochastic grammar	Rectangular regions	Images of documents	Not shown	Images with the results
[Peng et al. 2008]	Others	b	AND-OR graphs	Templates of atomic parts	Lotus Hill Institute	Not shown	A construction of an image database
[Glomb 2007]	Others	a	Sparse kernel feature analysis, EDT graphs	Set of pixels	Cars, images captured in urban environment	Generation of EDT graphs	Presentation of the generated graphs

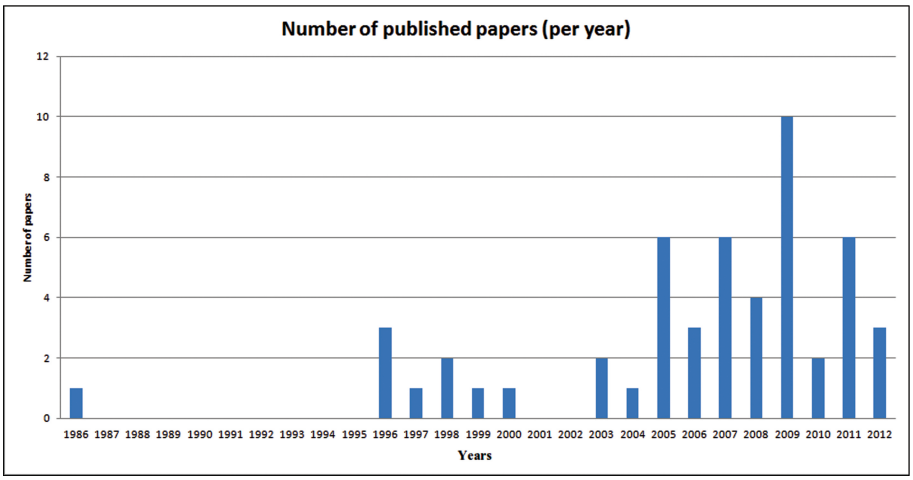

Fig. 1. Number of papers published every year.

many different images, (ii) more realistic interface models to map primitives to the pixels in the image, (iii) more powerful algorithms for discriminative classification and generative methods, and (iv) large volumes of digital images for training and testing.

Figure 1 shows that there were publication peaks in 2005, 2007, 2009, and 2011. In 2005, the main techniques used in the four articles were AND-OR graphs, adjacency grammars, and 2D tabled matrix grammar. In 2007, the techniques used were spatial random tree grammars, fuzzy grammars for texture recognition, EDT graphs, and the L-system. The year 2009 had the highest number of published articles (10 papers). AND-OR graphs appeared in five papers. The others showed context-free grammars for pattern recognition in medical imaging, fuzzy grammars for pattern recognition in texture patterns, and L-system grammars for object construction. In 2011, the main techniques used were AND-OR graphs and grammar, context-sensitive stochastic grammar, models of uncertainty ETPL(k) grammatical graphs, and shock graphs and grammar.

Table I shows that several types of grammars were used and that they are usually combined with various other techniques for pattern recognition. Although all the articles analyzed used grammars, we found different purposes for their use: some used them to recognize objects [Han and Zhu 2005]; others focused on the recognition of patterns into textures (e.g., Ferreira et al. [2009]); some used syntactic methods for creating new objects, usually trees and flowers [Schlecht et al. 2007]; in Wu and Bian [2009], grammars were used in the task of image segmentation, and Wang et al. [2005a] presented a method for working on image scaling; some papers deal with the problem of document layout recognition [Kanungo and Mao 2003]; and some papers addressed the problem of image storage [Peng et al. 2008] and how to convert an image into a set of terminal symbols [Glomb 2007]. Figure 2 displays a graph that illustrates that syntactic methods were the most used for object recognition (63%) in comparison to texture pattern recognition (6%), creation of new objects (13%), image segmentation (4%), scaling (2%), layout recognition (8%), and others (4%).

Figure 3 shows the main techniques found in the articles reviewed. We observed that 35% of the articles used some kind of stochastic grammars. AND-OR graphs were found in 16% of the studies. Fuzzy grammars and fuzzy logic appeared related to texture pattern recognition (6%). Articles showing L-system grammars working

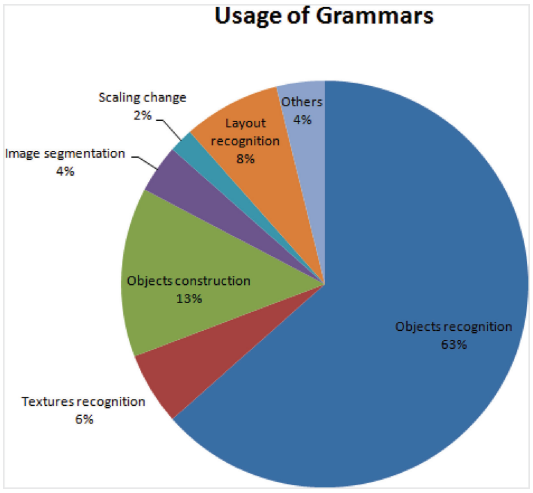

Fig. 2. Usage trends of syntactical methods.

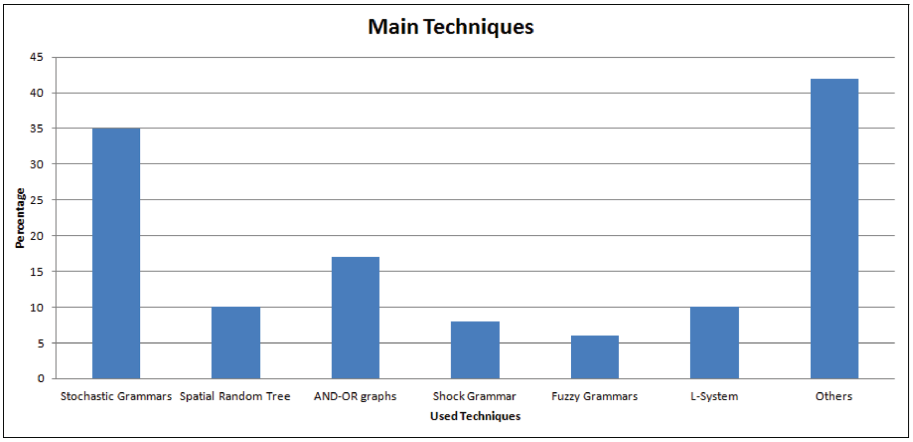

Fig. 3. Most used syntactic techniques.

on objects construction, especially modeling of trees, flowers, and fungi, represented 10%. Another technique found was the spatial random tree, which used the center surround algorithm for classifying objects (10%). Shock graphs and shock grammars were used for object recognition in 7% of the papers analyzed. Articles with none of the aforementioned techniques represented 42% of the studies analyzed during this systematic review.

Table I shows the strong presence of statistical methods used in the task of pattern recognition. The most common statistical methods were the hidden tree models, mixture of hidden Markov models, and Markov chain Monte Carlo.

Both qualitative and quantitative assessments were commonly found in the papers analyzed. For the qualitative analyses, the images obtained by applying the technique or model are always displayed, while for the quantitative analyses, true classification and misclassification rates are usually displayed, as well as ROC curves.

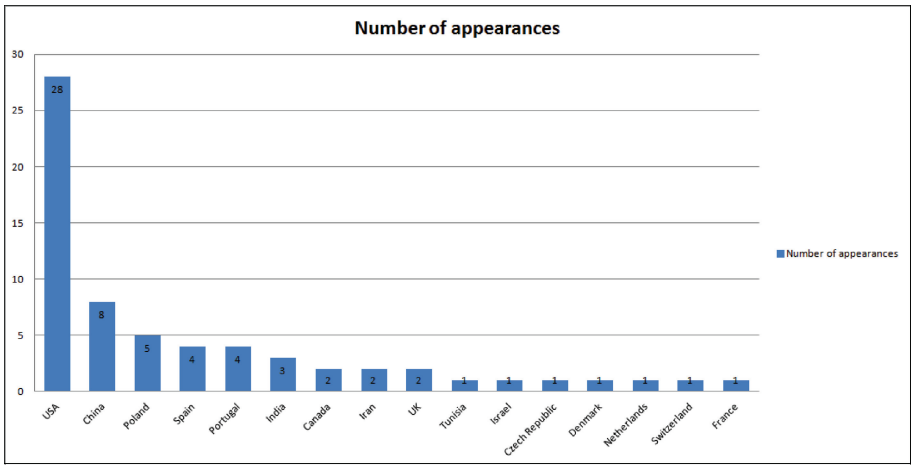

Fig. 4. Number of publications for each country.

Figure 4 illustrates that the United States was the country with the highest number of publications (28), followed by China (eight), Poland (five), and Spain and Portugal (four).

3.2. Object Recognition

3.2.1. Stochastic Grammars. The article by Stuckelberg and Doermann [1999] presented a probabilistic approach to image decoding. The approach seeks to maximize the posteriori of an objective function that represents the document and uses a task-oriented model to decode the document. Object recognition and structured image interpretation are performed by a single engine, allowing the integration of a noise model within the optimization process. The model used extends a stochastic context-free attribute grammar by introducing priors on the distributions of attribute values and measures and models for all symbols of the grammar (terminal and nonterminal symbols). In this model, a scanned page is equivalent to the initial symbol of the grammar. Next, some refinements are performed (horizontal and vertical cuts) in order to locate the *score page*, the *staff systems*, the *staves*, the *symbol groups,* and so forth. Each object is associated with a set of measurements that can be performed on the image and a set of models (such as hidden Markov models) that predict the outcome of these measures as a function of the object parameters. The approach is segmentation-free and separates the document model from the recognition algorithms. The article does not present results and conclusions with regard to the approach used.

Spatial Random Tree Grammars

Spatial Random Trees (SRT) were first proposed in Wang et al. [2005b, 2006]. SRTs are stochastic hidden tree models where leaf nodes represent parts of an image. In this model, the states at the tree nodes are random variables, with its whole structure randomly generated by a probabilistic grammar.

Multitree dictionaries are used for the SRT model development and defined using the formalism of grammars. The symbols of this grammar can represent different regions of an image. The production rules can divide these regions into smaller regions by recursive applications.

The SRT construction requires an image representing a domain Q and a grammar G that describes the hierarchical segmentation of the domain Q. The initial symbol of G

is the tree root, the nonterminal symbols will be the internal nodes, while the terminal symbols appear only in the leaves. In addition:

—the initial symbol is given by $a = (0, Q)$;
—the nonterminal symbols have the form $a = (j, R)$, where R is a subset region of Q (obtained by subdividing Q) such that $|R| > 1$ and j is an integer from a fixed set $\{1, \ldots, J\}$, which may represent, for example, the classification of the image pixels belonging to R in one of the J classes;
—the terminal symbols have the form $a = \{j, \{n\}\}$, where $\{n\}$ is a 1×1 rectangle, a pixel ($n \in Q$), and j is defined as above.
—the root production is $(0, Q) \rightarrow \{(j, Q)\}$ for $j = 1, \ldots, J$;
—the other rules have the form $a \rightarrow \alpha$, where $a = (j, R)$ is a nonterminal and $\alpha = (j_1, R_1), \ldots (j_{|\alpha|}, R_{|\alpha|})$ such that:
  —(j_k, R_k) is a terminal or nonterminal symbol;
  —$R_1, \ldots, R_{|\alpha|}$ are disjoint regions that partition R.

Finally, a probability distribution is defined for each nonterminal present in the tree. The authors also proposed an exact inference algorithm of polynomial complexity for parameter estimation, as well as algorithms for other inference problems involving SRTs. Another important contribution is the development of algorithms that can be used in the tasks of segmentation and classification. The true classification rate of the proposed model was of approximately 97% when images of houses, buildings, and shop fronts were analyzed.

This framework evolved into a more flexible version, described in Siskind et al. [2007], which can be applied not only to images but also to any other multidimensional data. In the new formulation, a terminal symbol is a feature vector that, in the case of images, is extracted from an image region obtained through a segmentation algorithm.

AND-OR Graphs

An AND-OR graph is a six-tuple that represents an image grammar $G = <S, V_N, V_T, R, \sum, P>$, where S is the initial symbol representing the scene or the category of an object, V_N is the set of nonterminal nodes, V_T is the set of terminal nodes (primitives, parts, and objects), R is the set of relationships between nodes, $\sum$ is the set of all valid configurations that can be derived from G, and P represents the probability model defined in the AND-OR graph. To represent context information, the AND-OR tree is transformed into an AND-OR graph by adding horizontal edges connecting nodes that have some type of relationship (OR nodes that are children of an AND node). The addition of these edges produces an AND-OR graph representing a grammar of the image. A syntactic graph is a parse tree of this grammar with the increment of the relationships between nodes. The stochastic model of an AND-OR graph must define probabilities in the syntactic graphs [Zhu and Mumford 2006].

A top-down/bottom-up algorithm is used to transverse scroll through the AND-OR graphs. It is a greedy algorithm to maximize the posteriori Bayesian probability that is processed in three phases. The first phase is the bottom-up detection, in which hypotheses are created with a certain weight. In the second phase, the algorithm chooses the hypothesis with the greatest weight among all candidates, and if accepted it increases the Bayesian probability. In the third phase there is the top-down/bottom-up integration. Each hypothesis in the current syntactic graph matches a production rule with attributes passed to the nonterminal node. These nonterminal nodes also match other production rules, which then generate top-down proposals for prediction. The top/down proposal weights are calculated based on the posterior probabilities. Thus, whenever a new node is added to the syntactic graph, a new subconfiguration is created, enabling the following actions: creation of potential new top-down proposals, inserting them on

the list; redefinition of weights of some candidate proposals; and transmission of attributes between nodes and their parents by means of restriction equations associated with the production rule. At the end of the algorithm a configuration that represents an image is obtained.

The usage of AND-OR graphs in computer vision was first introduced in Chen et al. [2006]. In this paper, these graphs were used to deal with a large variety of cloth configurations, for example, T-shirts, jackets, and so on. First, an artist was asked to draw sketches of dressed people. Next, these sketches were decomposed into clothing categories, such as collars, sleeves, cuffs, pants, and so forth. Each one of these categories was represented by a subgraph that was used as leaf nodes of a large AND-OR graph. In the AND-OR graph an AND node is used to represent a decomposition of the graph considering its subconfigurations, while an OR node is used as a switch that allows one to choose one of the AND node alternatives. The structure proposed integrates the stochastic context-free grammar to represent structural variability using Markov models for context. The results of the use of AND-OR graphs for clothing configurations can be seen in images in Chen et al. [2006].

In the paper by Zhu et al. [2009], AND-OR graphs were used for object recognition in combination with oriented triplet primitives. These oriented triplets are formed by three points of interest, each point represented by its location, orientation, and appearance vector. These triplets have geometric properties that make them scale-rotation invariant [Zhu et al. 2009]. To create a theoretical framework for unsupervised learning of probabilistic models for the generation and interpretation of natural images, this paper presents the use of a model called Probabilistic Grammar Markov Model. This model combined Markov random fields and probabilistic context-free grammars. The task of model learning can be divided into two parts: learning the model structure and learning the model parameters. For the first task, an AND-OR graph is constructed iteratively as more triplets are added to the graph. For the second task, clustering techniques were used to propose adding new triangles to the graph and to validate or reject these proposals. The method presented is not dependent on position and scale object orientation. The tests were performed with images obtained from Caltech[13] (13 categories of objects) and the model presented 90% of true positives and 10% of false positives. These results are for ratings of images containing a searched object and images that only have the background.

In order to locate human faces in images, Reddy et al. [2009] proposed a method with a two-layer architecture. At the first layer, facial components are detected using a certainty factor-based geometrical model, while the second layer uses top-down/bottom-up algorithms to transverse an AND-OR graph to localize the face. The AND-OR graph combines a stochastic context-free grammar to represent object configuration variability and Markov random fields to represent the spatial relationship between the face components. The tests were performed with 3022 images, and the face detection rate obtained with the method was 93.2%, outperforming other methods based on neural and statistical approaches.

The papers by Han and Zhu [2005, 2009] showed an attribute context-sensitive grammar capable of representing an image. Moreover, these studies showed the effect of a top-down/bottom-up algorithm used for parsing in the process to maximize the posteriori probability. The terminal nodes of the grammar created are planar rectangles projected on the images. All nonterminal and terminal nodes are described by attributes that represent their geometric properties and appearance. The grammar submitted has six production rules. A rule expands the root node (scene) in m independent objects. Another rule instantiates a nonterminal node in a primitive. The

[13]http://www.vision.caltech.edu/html-files/archive.html.

remaining four rules arrange objects or surfaces recursively into four possibilities: (i) align m objects in one line, (ii) nest an object inside another, (iii) align three rectangles to form a cube, and (iv) arrange objects in the format mxn in a mesh. Given an input image, the goal is to compute the parse tree in which each nonterminal node corresponds to a production rule.

Rothrock and Zhu [2011] presented an AND-OR graph grammar to represent a body as a set of articulated compositions and reconfigurable parts. The fact that the model is reconfigurable allows replacing compatible parts with different attributes such as clothing appearance. Each production rule defines an appearance model for the root part, a set of child parts, and a set of constraints over the geometries and attributes of the child parts. A probabilistic model is defined for the parses and exact inference is computed for certain restrictions of these constraints. The results were competitive with the state-of-the-art data: 88.4% (method presented) against 86.9% (state-of-the-art) of correctly estimated parts on average.

In Parag et al. [2012], a meta-grammar used to write rules for object detection and recognition was submitted. The object description rules are oriented toward an AND-OR structure. In the proposed grammar, the set of terminal symbols are formed by low-level image features. The object representation is constrained to be hierarchically expressed, which allows accommodating the construction diversity. To substantiate the proposed grammar for a specific application case, a compiler used to parse and translate an object grammar into an implementation in PROLOG of a BLR[14] logic program was implemented. The technique was used to detect graphical user interface (GUI) components for component detection and automated software testing processes. A set of six images was tested, with a total of 20 to 40 objects per category. The detection rate was 90% for push button, 100% for menu list, 90% for radio button, and 67% for checkbox, and the number of false alarms were 0.17, 0, 0.33, and 2.67, respectively.

In Si and Zhu [2011], a framework that uses a model called AND-OR Template (AOT) was presented. In this model, an AND node represents hierarchical composition, and an OR node represents deformations of parts. This paper shows that not only the structure but also the parameters of the AOT can be learned from images in an unsupervised way. The learning process is composed by a block-pursuit procedure used to learn the dictionary of primitives that form the leaf nodes, AND nodes, and structural OR nodes. Furthermore, a graph-compression operation is used to improve generalization; this process generates new OR nodes in the compositional hierarchy. Two different tests were promoted in order to test the proposed framework. The first one used 1D examples (words), and the idea is to study the factors responsible for model identifiability (*n:* training sample size; *s:* used to control the average length of random letters added between two words; α : a parameter used to control the compression of the graphs). For $n = 100$, the framework presented sensibility $= 1$ when $s = 0$ and the false-positive rate is about 0.068; for $n = 100$, the framework presented a sensibility $= 1$ when $s = 1$ and the false-positive rate is about 0.064; and for $n = 100$, the framework presented a sensibility $= 1$ when $s = 2$ and the false-positive rate is about 0.055. These results were obtained comparing the real dictionary of words with the learned one. The other test was for object recognition (egret, deer, and bikes), which used close to 20 images for training and a large number of images for testing. The results of this test were compared with Latent Support Vector Machine (LSVM), based on the recognition of objects, parts, and key points. The OAT framework demonstrated a better performance when compared with LSVM.

[14]BLR stands for bilattice-based logical reasoning.

Stochastic Context-Sensitive Grammars

Jin and Geman [2006] proposed a mathematical framework to construct a probabilistic hierarchical image model that can use arbitrary contextual relationships. Different from a traditional formal grammar, the proposed architecture allows the sharing of subparts among many entities. In addition, it does not limit interpretations to single trees (i.e, zero or more instances of the modeled pattern that can be detected). The proposed method was tested in a set of 385 images containing license plates, and the detection rate of plates was above 98%.

Tylecek and Sara [2011] use stochastic context-sensitive grammars to recognize regular structures exhibiting symmetries. It uses the concept of scopes, where a configuration of objects is placed inside a "container" that is made up of images (a rectangle, for example). Scopes can represent, for instance, arrays of windows. In addition, spatial and structural exceptions were introduced (presence of holes) into otherwise regular arrays of elements. The idea is to depict a semiregular array of elements. A lattice is the spatial layout of the terminal elements with possible individual spatial deviations of locations. Tylecek and Sara [2011] represented this structure using an attributed context-sensitive stochastic grammar. The grammar describes how an image contains scopes.

Wang and Jiang [2009] presented a framework that uses a context-sensitive graph grammar to extract roofs of buildings displayed in aerial images. The proposed grammar is composed of five production rules and three types of shared quadrilateral primitives. Each production rule contains equations to constrain the attributes of a parent node and those of its children. Furthermore, links between nodes at the same level were created to represent their spatial and appearance relationships. To extract the roof, a bottom-up algorithm is used, which generates hypotheses (quadrilateral) by grouping structures viewed in the image, such as straight lines, parallel lines, and so on. The authors did not present experimental quantitative results.

3.2.2. Markov Models and Fuzzy Logic. Soltanpour and Ebrahimnezhad [2010] used a structural context descriptor for extracting objects in images by recognizing parts of objects with similar structures. Blocks from oriented triangles are used to generate a grammar from the parts of objects, where each block is constructed from the central parts of such objects. This structure increase is based on the number of parts detected by adding new triangles. The model is trained using Mixture Hidden Markov Model (MHMM) and an Adaptive Neuro-Fuzzy Inference System (ANFIS) for object categorization. To test the approach, the proposed model was used in four image classes: motorbikes, faces, planes, and cars. Using ANFIS, recognition rates were of 75.3%, 88%, 85%, and 70.2% for parts of each of the classes, respectively. Using the MHMM, the recognition rates were of 96.2%, 98.3%, 95.2%, and 82.7%, respectively.

3.2.3. Augmented Regular Expressions. Augmented Regular Expressions (AREs) are regular expressions with the addition of a set of constraints that involve the number of instances of the operands of the star operations[15] in each language string. Such constraints increase the descriptive power of regular expressions, enabling the description, recognition, and learning of a nontrivial class of context-sensitive languages [Sainz and Sanfeliu 1996]. Sainz and Sanfeliu [1996] described a method to learn the grammar model from a set of positive and negative samples. Each model is represented by a pseudo-dimensional ARE, where each line is represented by an ARE and all columns together are represented by another ARE. The paper describes the learning process and the results of applying this method to learn traffic signs. During this learning

[15]In regular expression, if V is a language, then V-star can be described as a set of all elements that can be formed by concatenating of zero or more V elements.

process, image model lines are represented by a finite state automaton (FSA). The method used for FSA learning is called *active grammatical inference*. The process is conceived as a sequence of learning cycles, and each cycle includes a combination of symbolic and neural techniques, where the control of the next neural training is dynamically modified by the information acquisition or by the imposition of external information.

3.2.4. Predicate Logic. Shet et al. [2009] presented an approach based on predicate logic to recognize and detect patterns of objects in images in which two classes of objects are explored: pedestrians, watched by security cameras, and complex man-created structures that can be seen in satellite images (*surface-to-air missile,* SAM). This method analyzes a pattern of objects through modeling and specification of grammars. These grammars are encoded as first-order rules and the object analysis corresponds to the search through the space feature for the best solution, satisfying the logical constraints. For the problem of human body parts detection, a cascade of Support Vector Machine (SVM) classifiers was trained on histograms of gradient orientations, and a sigmoid function was applied to the output of each SVM. The rules to validate or reject human hypotheses are based on geometric information entered a priori in the system, such as the expected height of people and the expected base location of regions.

3.2.5. Adjacency Grammars. An adjacency grammar is a five-tuple $G = (V_T, V_N, S, P, C)$ where V_T is the set of terminal symbols, V_N denotes the nonterminal symbols, $S \in V_N$ is the start symbol, C is the set of constraints applied to the grammar elements, and, finally, P represents the set of productions defined as $\alpha \rightarrow \beta_1, \ldots, \beta_n$ if se Γ ($\beta_1, \ldots, \beta_n$) where $\alpha \in V_N$ and all $\beta_j \in V_T \cup V_N$ and Γ is an adjacency constraint defined in the attributes of β_j.

The symbols β_j can appear in any order. For example, for the production rule $\alpha \rightarrow \mu, \nu, \sigma \in P$ should be considered all the six possible symbol permutations of μ, ν, and σ as an equally valid replacement for the symbol α.

Studies on graphic design recognition, mainly focusing on the recognition of hand-drawings and gestures, are presented in Mas et al. [2005, 2008]. The recognition process shown consists of three main phases. The first phase, primitive extractions, consists of approximating drawn traces by primitives (arcs and line segments). The second stage is a syntactic stage for the recognition of compound objects where symbols of a given class are recognized by an adjacency grammar. The last phase is responsible for applying semantic rules to the instances of recognized symbols in the drawing. The created approach was tested in more than 700 symbols and obtained a recognition rate above 87.7% [Mas et al. 2005].

3.2.6. ETPL(k) Grammar Graphs. The articles by Trzupek et al. [2009, 2011] presented an approach for the interpretation of medical images. In particular, the articles focused on the recognition of a heart disease (stenosis) that can be detected from the 3D images of coronary arteries. After selecting the image to be analyzed, the system executes a skeletonization process that represents its morphological structure. The skeletonized image is used to form a graph where each vertex represents a branch point of the main artery, and this graph represents the spatial relationships between arteries. This spatial representation can be formalized by using graph grammars called *Embedding Transformation-preserved Production-ordered K-left nodes unambiguous* (ETPL(k)). According to the authors, the system is able to recognize locations, amounts, and types (concentric and eccentric) of stenosis with an accuracy classification rate of 85%.

3.2.7. Context-Free Grammars. The approach presented in Ogiela et al. [2008] was established for pattern recognition in human bone radiographs, specifically bone fractures and fissures of arms and legs. The created context-free grammar can recognize the

following types of fractures: fissures, spiral, longitudinal, displaced fracture, delayed, transversal bond, and adhesion. In this grammar, the terminal nodes (*a, b, c, d, e, f, g, h*) represent angle intervals. For example, the terminal node *a* represents the interval between –10 and 10 degrees. The proposed grammar was able to recognize the angle between the parts constituting the human bones in order to identify the fractures. The created system achieved more than 90% performance accuracy in the interpretation of bone fractures of human legs.

A method for indexing and recovering images was presented in Ogiela et al. [2009]. The presented context-free grammars are able to recognize the topology of organs and body parts (arteries, bones, and pancreas), enabling one to identify visible injuries and diseases in these organs and structures. One of the used techniques creates a grammar based on the syntactical graphs of the segmented images (obtained by X-rays) of human bones. In this technique, terminal symbols are represented by the gravity centers of each hand bone. Another technique uses a grammar based on the arterial topology and pancreas structure, where it is possible to model arterial stenosis and assist in cancer diagnosis. The application of the proposed techniques attained a recognition rate of about 93% when used to analyze and interpret lesions.

Gidas and Zelic [1997] developed an object recognition method that uses the advantage of bottom-up and top-down methodologies. The approach explores conceptual aspects from context-free grammars, nonparametric statistics, and polygonal fields. The approach is formed not only for a Hierarchical Syntactic Models (HSM) used for contextual constraints, such as decomposition of objects into parts and subparts, but also for data models used to depict likelihood functions of the observed data considering each instance of the object representation. The data models, described by non-parametric statistics, are invariant under imaging conditions (e.g., contrast). In an HSM, types of objects are viewed as junctions of articulated joints and parts; for example, a car has doors with windows, but also it has lights, license plates, and so on. Generally an object is a concatenation of primitives. The tests were performed in a database of simulated tools, like hammers, shovels, scissors, and so forth.

Toshev et al. [2010] presented a framework to build the symbolic representation of detection and parsing. In this framework, an unorganized point cloud is used as input as result and a parse tree is obtained whose nodes are surfaces and volumetric parts. The volumetric parts were considered to increase robustness against occlusion, which can be inferred from subsets of their bounding surfaces. The idea is to see a building as a tree, whose nodes represent volumetric parts that are next to each other and covered by roofs (planar patches), which are considered children of the volumes. To do this, a simple grammar was introduced to capture geometric properties between volumes and planar patches. The proposed grammar has two supernodes, "building," used as ascendants of all other nonroot nodes. This feature allows performing detection while parsing, where all parse trees rooted in the "building" node are considered building representations. Dependency parsing, known as an efficient parsing technique, is used to infer the parse trees of buildings. Furthermore, labeled data can be used to estimate grammar parameters using structured learning. Tests were performed using The Wright State 100 dataset, containing approximately 1 billion points collected by airborne and terrestrial range scanners. The points were divided into 350 blocks. The accuracy achieved was of 89.3%, against 87.9% achieved by a support vector machine method.

3.2.8. Regular Grammars. A method to interpret and describe how the trajectory of an object can be modeled using qualitative features was presented in Christensen et al. [1996]. The authors stated that the object evolution in a scene can be described by regular grammar rules that must be based on observable characteristics. Moreover, it must describe individual objects and their relationships, as well as include action

compositions. Examples of characteristics include recognition of objects (cup, saucer), geometric relationships (aligned, parallel), temporal discontinuities (entering a view field, static), and so forth. When one of these characteristics is found in a particular state, a particular production rule is activated representing a transition state. In addition, each transition is associated with a specific action (handling, initialization, termination, etc).

The system was tested with a sequence of images depicting a breakfast table. The scene description was performed by a generic parser with 42 production rules. The analysis result is the following textual scene description: a saucer was placed on the table; a cup was placed on the table; a cup with a saucer was placed on the table; a spoon was placed on the table; the teapot is now on the table; the milk jug is on the table; the table is set, 'Tea is now served!'.

3.2.9. Three-Dimensional Plex Grammar. Conventional string grammars consider that each symbol, terminal or nonterminal, has two attaching points: on the left and on the right. Plex grammars generalize it, considering that symbols may have n attaching points [Lin and Fu 1986]. These are special cases of attribute grammars, where the connections between symbols are the attributes of each production.

The paper by Lin and Fu [1986] created an extension of plex grammars for the recognition of three-dimensional objects, called three-dimensional plex grammars. In these grammars, the symbols are primitives or composite surfaces having n attaching curves to join into other surfaces, which are called *n attaching-curve entities* (NACEs). One component of this grammar is a set of identifiers that describe the NACEs attaching curves.

A plex three-dimensional context-free grammar has productions in the form $A\Delta_A \rightarrow \chi\Gamma_\chi\Delta_\chi$, where A is a single NACE, Δ_A is the list of curves on the left side (*left-side tie-curve list*), χ is the list of NACEs on the right side, Γ_χ is the list of interconnections on the right, and Δ_χ is the list of curves on the right side (*right-side tie-curve list*). The list of NACEs on the right side is a string in the form $\chi = \mathrm{a}_1\mathrm{a}_2 \ldots \mathrm{a}_j \ldots \mathrm{a}_n$, where a_i is a single NACE. Γ_χ specifies how and which NACEs are interconnected.

A system that uses these grammars comprises two main steps: analysis and recognition. The analysis consists of the selection of primitives (surface fragments) from a 3D model object and creating a three-dimensional plex grammar to represent it. The recognition step uses this grammar to recognize any projection from the 3D model in input 2D images.

3.2.10. Structural-Syntactic Approach. The study by Gao et al. [2000] presented a proposal based on a combination of statistical and structural methods for recognizing Chinese characters. In the structural approach used, a Chinese character is represented by a hierarchical structure where the primitives are formed by the features used in the drawings. The statistical part of the proposed framework is responsible for comparing the image features being analyzed with the information obtained from the training examples. The prototype built was trained with 7920 samples of handwritten Chinese characters and returns a score in the range [0, 1] as response of the image analyzed against all the examples from the database.

3.2.11. Shock Grammar. A 2D representation of an object that preserves the silhouette of an object is called a skeleton. These skeletons can be converted into a shock graph, which is a shape abstraction used to group skeleton points according to the variation of a radius function. These groups are called "shock groups" and they are labeled from 1 to 4. Such a grouping decomposes a skeleton into parts, and the interrelation confirms a well-defined grammar. A shock graph is obtained from the shock groups and is a directed acyclic graph [Hingway and Bhurchandi 2011].

Zaboli and Rahmati [2007] used shock graphs, a skeleton-based method, for object recognition. The method uses branch points (which have three or more neighbor points in their eight-neighbor pixels) in shock graphs and shock grammars in order to include these features to improve the recognition rate and efficiency in object recognition. The *medial axis* was used to retrieve the skeleton from an image. The medial axis of a shape is considered the central point of the largest circles contained within the curve, touching the boundary of the shape. The branch points are considered the basic points of the skeleton of a shape. Generally, these points are the internal parts of the skeleton and are robust under deformations. Furthermore, these points are, generally, invariant in the presence of missing parts and occlusion. The matching trees used in Zaboli and Rahmati [2007] were not based only on their topologies, but they also considered geometric information for each vertex and its corresponding points in the skeleton. The tests were performed in order to determine the similarity between two given shapes. The similarity is normalized with values between [0, 1], considering that the highest similarity produces values closer to 0.

In the method proposed by Hingway and Bhurchandi [2011], every object in the database is converted into a binary image, hence obtaining its skeletons. These skeletons are converted into shock graphs. The shock graph of a query image is compared to the shock graphs of the database images. The algorithm to compare the shock graphs of two images is as follows: (i) split the graph for any object and the graph for the query image into subgraphs; (ii) compare the subgraphs of the object to that of the query image; and (iii) when comparing the subtrees, compare the type of shock and, if the type does not match, consider it a mismatch. The sum of all matching scores is achieved by the total matching score.

Siddiqi and Kimia [1996] developed a theory for the generic representation of two-dimensional shapes, where structural descriptions are derived from the shocks (singularities) of a curve evolution process, acting on bounding contours. The approach used to classify a shock is based on differential properties of an *embedding surface*. This idea was for developing an implementation to achieve accurate geometric estimates in the neighborhood of discontinuities in order to localize shocks. The constraints used to prune unfeasible shock configurations are defined by a shock grammar. As the defined grammar does not describe topological and geometric constraints, the grammar is embedded in a graph.

In Siddiqi et al. [1998], the researchers applied the theory to the problem of shape matching. The shocks were organized in a directed, acyclic shock graph. The spacing of the graphs is characterized by the rules of a shock graph grammar, which allows reducing a shock graph to a unique rooted shock tree. Thus, the authors used a tree matching algorithm, which finds the best set of corresponding nodes between two shock trees in polynomial time. They demonstrate the system's performance by using a large database of varied shapes that take into account articulation, occlusion, and viewpoint changes.

3.3. Texture Recognition

The studies that have focused on the pattern and texture recognition analyzed in this review used fuzzy grammars and belong to the same group of researchers.

The papers by Ferreira et al. [2007a, 2007b, 2009] presented a system for monitoring cork pieces, which used a texture segmentation approach based on the wavelet transform and a fuzzy grammar as classifier. The features extracted from the images were processed with the wavelet transform, with this technique applied to each color component in order to perform a color-based texture analysis. The developed approach is formed by the learning and monitoring phases. In the learning phase, the monitoring texture is manually selected by specifying a region of interest in the image. A feature vector is extracted and a fuzzy rule to characterize the texture is used to generate

the fuzzy grammar. In the monitoring phase, a feature vector is extracted from the image under analysis. The fuzzy grammar generated in the learning stage is then used to evaluate this new feature vector and perform the classification. According to the authors, an advantage of this approach is that the system can be trained with only one texture sample. Furthermore, the system can be used in environments with many types of textures, even if the lighting conditions are unstable. The feasibility and efficiency of the approach have been proven in experiments with more than 30 types of textures extracted from images with 640×480 pixels. The system responded with a value higher than 0.85 for the rule describing the texture used as input, and a value lower than 0.3 for rules corresponding to the other textures. A higher value indicates a greater possibility of the texture being described by the rule used.

3.4. Object Construction

3.4.1. Attribute Grammar. The paper by Luo et al. [2009] presented a study on the perception of three-dimensional objects from two-dimensional sketches manually annotated from an image. The attribute grammar used is defined as a quadruple $G = (V_N, V_T, PR, R)$. V_N and V_T are the sets of nonterminal (2D aspects/3D parts) and terminal nodes (2D elements), respectively. PR is the set of production rules used to build pyramids, cubes, prisms, and so forth. R is the set of relationships between each pair of nonterminal nodes. The main idea of this project was to recognize objects from a 2D image and create a new version of these recognized objects in 3D. The approach used is composed of two layers, one 2D and the other 3D, which together compose a hierarchical grammar model for representing man-created objects. In the 2D layer, given an input image (2D sketch manually annotated), the model groups the geometric elements (triangles, rectangles, etc.), forming various 2D aspects through a top-down/bottom-up algorithm. Each one of these aspects is used to make inferences about hidden structures of the 3D part. These two layers are calculated recursively in a Bayesian framework using data-driven Markov chain Monte Carlo. The tests were carried out with eight different types of objects using 20 images for each category. The time for each analysis was approximately 60 seconds. For each object category modeled by the system, another object was created in 3D using computer-aided design (CAD). All objects were mixed, and then individuals were asked to point out which object had been created by the proposed method and which had been created using CAD. An accuracy classification rate equal to 50% was expected, since the idea is that the individuals would be unable to differentiate an object created using the aforementioned technique from an object created using a CAD, and the observed rate was 52.85%.

3.4.2. Matrix Grammars. The paper by Subramanian et al. [2005] provided a system for generating rectangular figures using a system called *splicing array grammar systems* (SAGS). This paper reports one the formal specifications of a 2D tabled matrix grammar, the operations that can be performed with these arrays (operations on the rows and columns), and the system SAGS itself. The images formed by the proposed method consist of letters in a matrix representation.

3.4.3. L-System. Given a parallel grammar $G\| = (V_N, T_N, R, S)$, this grammar accepts the string x, if there is a sequence of parallel derivation $S \Rightarrow \mathrm{w}_0 \Rightarrow \mathrm{w}_1 \Rightarrow \cdots \Rightarrow \mathrm{w}_k = x$, where each derivation step ($\Rightarrow$) denotes the application of all possible rules for an intermediate string. The languages accepted by the grammar $G\|$ and the corresponding nonparallel G is not the same, except for simple cases [Chanda and Dellaert 2004].

Parallel grammars have been used in applications of computer graphics in the form of L-system [Chanda and Dellaert 2004]. This type of system originated from a mathematical theory that explained the growth and development of multicellular organisms and had great impact on modeling and simulation areas (*apud* [Chanda and Dellaert 2004]). It is currently used in various computer graphics applications.

L-systems are parallel formal grammars, generally context sensitive. The production rules and symbols have their own semantic and visual interpretations based on the problems being modeled.

A stochastic version of L-systems was used in Schlecht et al. [2007] to represent fungal biological growth. The researchers presented a method for making inferences on the biological structure obtained from microscopy images. Considering that such inferences from a model are difficult due to the amount and independence of parameters, a sampling based on Markov chain Monte Carlo was created to explore the parameter space to search the parameters that probably generated the data. To test the proposed method, 184 images of the fungus Alternaria from two datasets were used. The sampling was performed using 10 random initial states in two sets, each one with 20,000 iterations.

L-system grammar types were also used in Prusinkiewicz et al. [1988]. This paper presented a modeling method to simulate the development and growth of plants. The approach has two main features. The first one is the ability to emphasize the time-space relationship between the constituent parts of a plant. For example, in various species some development stages can occur simultaneously. The second feature is the ability to simulate the plant growth, once the proposed approach is capable of representing a plant of different ages. To reproduce these two features, the proposed method used L-system grammars, as they have recursive production rules capable of simulating the development of plants.

In the paper by Sun et al. [2009], a method that combines rule-based techniques and images to create light 3D models of trees is presented. The proposed model can be divided into four steps: (i) restoration of a tree trunk structure from a 2D image, (ii) 3D reconstruction of the trunk skeleton using the binocular vision method, (iii) extraction of the axioms and production rules from the skeleton using 3D L-system grammars, and (iv) use of an interpretation algorithm of L-systems to create models that can be sent via the web to be rendered on the client machine.

The paper by Hemberg and OReilly [2004] presented an extension of the grammar used in the Genr8 system, which is used by architects to develop surface designs. A surface on Genr8 begins with a closed polygon and grows by means of simultaneous applications of the production rules. During the growth process, the surface is influenced by "external" factors, such as repellents, attractors, and borders. The Genr8 uses a grammar called *Hemberg extended map L-system*, which is a more complex L-system grammar. The developed system was used in the undergraduate course on Design Emergent Design Technologies of the Architectural Design Association - AA in London in 2004.

Qu et al. [2008] presented a rule-based grammar learning method to create an L-system grammar to model plant development. The initial step used image processing and pattern recognition techniques to retrieve the geometric and morphologic structures of the plant. Then, the data obtained in the first step were analyzed by hidden Markov tree and semi-Markov chain in order to compose bidimensional hierarchical automaton (BHA) parameters, which describe the plant ramification structure. The last step is to create an L-system grammar from the transformation of the BHA. To test the proposed method, two experiments were conducted using 35 images from leafless trees with 360° coverage around the plant. One of the limitations of the technique is the high cost to capture the tree structure from several images. Moreover, the structure from trees with leaves cannot be captured due to the branches' occlusion.

3.5. Image Segmentation

The paper by Wu and Bian [2009] presented a stochastic method for image segmentation. The approach used decomposes an image into its visual components and returns as output a hierarchical representation in the form of an AND-OR graph. The problem

of Bayesian inference is first formulated, and then the solution space is decomposed into a union of various subspaces of varying dimensions. The goal is to optimize a posteriori probability. Subsequently, a top-down approach is used to describe how objects and region models (texture, shading, etc.) generate the image intensity. Finally, to perform the parameter estimates, bottom-up proposals are conducted to guide the search through the parameter space.

In Hamdi et al. [2012], a method to segment a given image and to estimate the area using regular grammars is presented. An image is defined as a set of words based on an alphabet, and an object, similar to a word, is recognized by an automaton. The image is divided into small blocks, each one having a predefined structure that represents the terminal node of a grammar. After, each block is associated to a number that performs the mapping of the represented structure in such a way as to create a numeric matrix, which represents the image. The next step creates a regular expression representing a closed contour. The automaton equivalent to this regular expression is used to analyze the matrix in order to search submatrices with these contours, performing the image segmentation. The researchers conducted tests with cardiac images and they indicate an accuracy of 93.22% to estimate the segmented areas.

3.6. Change of Scales

A study on perceived scale of space by increasing the scale from a traditional image is reported in Wang et al. [2005a]. The approach uses representation of primitive sketches that divide the image into structural parts (sketchable) to delimit the object's boundary and the texture parts (nonsketch). The approach also uses a type of graph grammar called *sketch pyramid* to represent topological changes. To construct these graphs, the proposed method uses a Bayesian framework and Markov chain Monte Carlo reversible jumps. The technique used was created with two different proposals. The first one, multiscale object tracking, is used to perform the monitoring of objects in a scene. The second proposal is called adaptive image display and addresses the problem of displaying high-resolution images on low-resolution displays (phones, PDAs etc.) and makes use of quadtrees.

3.7. Layout Recognition

3.7.1. Regular Grammars. Kanungo and Mao [2003] use stochastic regular grammars to represent document layouts as a hierarchical structure; each node of this hierarchy describes a document region. The hierarchy and the language are defined by the user, whereas the probabilistic parameters are estimated from training samples. The physical layout structure of a document is recognized through a parsing-based segmentation algorithm. More robust results, considering noise document, were obtained using a model with a set of estimated state duration distributions to incorporate the information on the physical style layout parameters.

3.7.2. Context-Free Grammars. An approach to hierarchical segmentation and layout structure recognition of the documents by using stochastic context-free grammars is presented in Shilman et al. [2005]. Machine learning is used to select the characteristics and all parameters used in the syntactical analysis process. Dynamic programming was also used to find the best derivation tree for a page. Two case studies were used to test the approach. In the first one, a generated framework uses a grammar created manually to analyze the document layout, obtaining an F-measure[16] of 85%. In the second case,

[16]F-measure is the harmonic mean of precision and recall.

the framework used an inferred grammar in order to support math equations found in the document, attaining an F-measure of 94%.

3.7.3. Context-Sensitive Grammars. Kong et al. [2012] use spatial context-sensitive graph grammars to recover the semantic structure in a web interface. The grammar performs the semantic grouping and interpretation of segmented objects from a screenshot image. A spatial graph is built where each node represents a recognized object and the edges represent spatial relations between the objects (text, buttons, etc.). Next, a parsing is performed to discover the hierarchical structure involving these objects. The approach was tested in many web pages and the object recognition occurred in a satisfactory manner in less than 25 milliseconds. The approach is less complex, considering the amount of generated nodes, than an analysis based on the HTML source code.

Mao et al. [2003] presented a technique aimed at particularly detecting technical paper title pages. To do this, hidden semi-Markov models and a stochastic attributed *K-d* tree grammar were used. The authors regard document analysis as a syntactic analysis problem, since the order of the components in a document page and their relations can be modeled as a grammar, which depicts the page at the component levels in terms of blocks and regions. The proposed model was divided into two parts: (i) a hidden semi-Markov model was used to describe the grouping of page regions into rectangular blocks; (ii) a *K-d* tree grammar was used to depict the hierarchical decomposition of the page. Initially, the document image is divided into parallel strips and then the number of black pixels in each strip is counted. Next, the sequence of pixels counted is used at a hidden semi-Markov model. The boundaries between different groups of strips are indicated by the state changes of the model, which labels the groups of strips. The vocabulary of grammar symbols proposed is formed by these labels, which is used to find physical layouts of the page.

3.8. Others

3.8.1. Image Storage, Indexing, and Retrieval. Peng et al. [2008] used AND-OR graphs to guide the image annotation in order to create an image database. The images are analyzed and their descriptions are transformed into AND-OR graphs. According to the authors, this type of graph allows precise data extraction. The database presented has two tiers: the first one manages the relationships of the visual concept model with AND-OR graphs, and the second one is used for parsing data based on physical objects in the database. The object category is used to index the AND-OR graph in tier 1. The data exported from the database can be in the form of raw images, object boundaries, interesting part patches, and so on.

3.8.2. EDT Graphs. A proposal consisting of deriving terminal symbols for some unknown class of objects to represent an image as a sequence of these symbols was presented in Glomb [2007]. To accomplish this task, the sparse kernel feature analysis technique is applied to a set of random fragment images in a training set, generating a sequence of sparse components representing the image characteristics. For each one of the analyzed images, the extracted features are used to derive the terminal symbol locations. Subsequently, an algorithm is applied to encode the terminal symbol positions to form a graph called *Extended Dominator Tree* (EDT), taking into account the symbol frequency and its configurations in the set of training images. The algorithm generates a sequence of symbols in the form of a graph that can be used for grammar rules learning methods. A potential limitation of this technique is that it may not be able to process complex objects, in which the spatial relationships cannot be adequately described. The tests were conducted using images of cars, and for each one of the 200 first images of the training set five fragments of 20×20 pixels were used.

4. DISCUSSION

Figure 3 shows the most frequently used methods/techniques in this research area: the stochastic grammars, AND-OR graphs, fuzzy logic and grammars, spatial random tree, shock grammar, and L-systems.

Stochastic grammars have been widely used, either traditional context-free or new formulations such as spatial random tree grammar, L-systems, or based on AND-OR graphs. Spatial random tree grammars were used for pattern recognition in images. The technique was used by only one research group, which proposed the model with an image recognition accuracy rate of approximately 97% [Siskind et al. 2007]. L-system grammars were the main approach used to construct objects, especially objects that can be constructed by means of recursive rules. AND-OR graphs were used by several research groups and were applied both in object recognition and in image segmentation tasks. For the latter problem, however, only a single article used this structure [Wu and Bian 2009]. Shock graphs and grammars appeared in four works. These techniques were used to recognize object shapes in binary images, as, for instance, in Siddiqi et al. [1998].

Grammar and fuzzy logic appeared in four studies. One of these studies [Soltanpour and Ebrahimnezhad 2010] addressed the problem of object recognition in images. The others are all from the same research group and addressed problems of texture recognition. It should be noted that all papers on texture recognition used grammars and fuzzy logic.

It was observed that there appears to be an unequal distribution in the papers concerning the application in computer vision problems. The vast majority of applications concentrated on problems of object recognition (33 articles out of 50), object construction (seven articles), layout recognition (four articles), and image segmentation (two articles). For each one of the other applications evaluated in this review (texture recognition and scale change), only one study was developed.[17]

We noted that most of the papers do not learn or extract the grammar structure from data, but rather estimate the probability distributions attributed to the production rules when the grammar is stochastic. Only the following studies perform grammatical learning rules: Zhu et al. [2009], which showed the construction of AND-OR graphs; Sainz and Sanfeliu [1996], which created augmented regular expressions with technical assistance from active grammatical inference; Ferreira et al. [2007a, 2007b, 2009], which built a fuzzy grammar for pattern recognition in textures; Sun et al. [2009] and Qu et al. [2008], which conducted learning of L-system grammars for tree constructions; Shilman et al. [2005], to recognize document layout; and Soltanpour and Ebrahimnezhad [2010], which used MHMM and ANFIS to learn a grammar.

Table II summarizes the main studies reviewed in this article. Although several studies use similar methods, they were selected for their innovative nature and because they presented further details on the techniques used. Moreover, all these papers presented either quantitative or qualitative results of the technique developed.

From the analysis conducted we observed the advantages and disadvantages of using grammars in computer vision. Furthermore, some gaps were perceived in the literature analyzed, which can direct toward additional opportunities of new research lines.

4.1. Advantages and Applications of Grammars in Computer Vision

The first advantage regards the more concise representation that grammars in general can provide, in comparison to the representation by an inherent set of pixels in the images. The syntactic approach can efficiently represent the pattern structures and

[17]The three articles dealing with the recognition of textures are about the same approach developed by the same research group.

Table II. Synthesis of the Main Papers Presented in This Systematic Review

Articles	Stochastic grammars	AND-OR graphs	Spatial random tree grammar	L-system	Fuzzy logic	Context-free grammars	Shock grammars
[Siskind et al. 2007]	X		X				
[Zhu et al. 2009]	X	X					
[Ogiela et al. 2009]						X	
[Prusinkiewicz et al. 1988]				X			
[Ferreira et al. 2009]					X		
[Han and Zhu 2009]	X	X					
[Siddiqi et al. 1998]							X

consequently facilitate the retrieval of images by means of their structures. Lin and Fu [1986] observed that the syntactic approach for pattern classification and scene analysis called for the researchers' attention, due to its abilities in handling pattern structures and their relationships. The syntactic approach is able to describe a large set of complex patterns using small sets of simple primitives and grammatical rules.

When images are perceived as structured scenes with a hierarchical order among the objects, their representation by grammars is still more intuitive. This is due to the fact that grammars can naturally describe hierarchical structures through their nonterminal and terminal symbols. This natural representation was emphasized by Christensen et al. [1996], who stated that the use of a grammatical approach in image context object recognition includes effortless recovering from errors.

Apart from other pattern recognition methods, a researcher interested in using a grammatical approach has many model options to choose, each one with a different representation power but also with a different complexity. Grammars of the most appropriate level in the Chomsky hierarchy can be chosen in order to achieve the expected results in the most efficient manner.

Most of the computer vision applications require grammars that are at least context-free. However, depending on the application goal, the highly efficient regular grammars can be used (linear time parsing). For instance, Kanungo and Mao [2003] used stochastic regular grammars to describe the layout structure of documents, and Hamdi et al. [2012] used regular grammars for image segmentation. Another advantage of using this type of grammars is the availability of efficient algorithms for grammar rules, allowing supervised learning of the used models, as performed in Kanungo and Mao [2003].

Context-free grammars demand polynomial time parsing algorithms, but they are able to characterize dependences between components at arbitrary distances in nested or branching structures. Such feature allows this type of grammar to represent richer structural patterns and to consequently be used in a broader variety of computer vision applications. For instance, context-free grammars can represent different lesions in medical images through rules of angle intervals between bones, vessels, and so forth [Ogiela et al. 2008, 2009]. Some learning algorithms of context-free rules are known [Takada 1988; Mäkinen 1992; Sakakibara 1992], for allowing to create the models from training samples.

Although regular and context-free grammars can be used, additional representation power may be necessary in order to better characterize complex visual patterns. Particularly, some forms of context information are important to represent the spatial relationship between image parts. In fact, most of the papers reporting success in complex object recognition have addressed this issue. Spatial random tree grammars augment stochastic context-free grammars by explicitly incorporating spatial information as part of the grammar [Wang et al. 2005b, 2006; Siskind et al. 2007]. Alternatively, stochastic context-free grammars can be combined with Markov random field, the former representing variability of the object configuration and the latter representing the spatial relationship between parts [Reddy et al. 2009; Zhu et al. 2009]. Spatial relationship can also be represented using attributed graph grammars [Han and Zhu 2005, 2009; Kong et al. 2012; Trzupek et al. 2009, 2011; Ogiela et al. 2009] or stochastic context-sensitive grammars [Yao et al. 2009; Tylecek and Sara 2011]. However, grammatical learning rules and parsing are problematic issues to be considered in the context-sensitive field. Fortunately, the constructions proposed in these papers are able to recognize a proper subset of context-sensitive languages keeping a polynomial time parsing. Parser generators for these grammar models could be very useful for new papers using these strategies. Grammatical learning rules for these proposed models, however, are addressed only in Zhu et al. [2009], where the grammar combined with Markov random field is context-free.

Alternative grammar formulations can also be explored. Fuzzy grammars were used in texture recognition and parallel grammars (L-systems) in object construction. L-system grammars can easily represent fractal patterns. Therefore, this model has been widely used to construct images such as plants, trees, and so forth, favoring the definition of models consistent with those used in developmental morphology and physiology [Prusinkiewicz et al. 1988].

4.2. Disadvantages of Using Grammars in Computer Vision

In spite of the concise representation provided by the syntactic representation, it is important to highlight some limitations of using grammars in the image context. Although there is a space economy in the representation, the processing computational cost can be high. As observed in this article, there are several methods, a dozen of objectives, and many types of possible representations. More complex patterns often require grammars with greater representation power and therefore with higher time complexity recognition. When the method requires a training phase, the processing can demand additional high computational time cost, depending on the size of the image sets and the type of grammar, among other factors.

A second limitation observed in most of the grammars used is the difficulty to process complex images. Here we consider as complex images those with high resolution and/or a high level of detail, and possibly including occlusions. Many of the studies presented in this article point out this limitation. For instance, when an algorithm is used to generate terminal symbols to represent images for EDT graphs, Glomb [2007] cited that his system may be unable to process complex objects. This author considers complex objects those in which the spatial component distribution is not sufficiently descriptive and/or there is a large number of visually similar parts.

Finally, other important imitations to be observed regard the high dependence of image preprocessing techniques that the grammars can present. Some of the papers mentioned in this review used techniques to execute operations to enhance structures, to smooth noises, and to segment structures. These techniques are well known in the literature of computer vision area as highly dependent on the image type and the processing objectives. Only a few authors discuss this issue, but failures in this previous process can lead to failure in the use of grammars, particularly when the goal is to establish procedures to store and retrieve features in order to classify images into categories.

4.3. Brief Comparison to Nongrammatical Techniques

Once the advantages and disadvantages of the general use of grammars are presented, it is important to delineate a comparison with other nongrammatical methods. In general, the use of grammars allows a more flexible way to represent images, mainly when the images have a well-established hierarchical pattern, as mentioned before. This occurs because it is possible to carry out a direct mapping between the image structure and the syntactic rules. Regarding this subject, Gao et al. [2000] stated that the use of structural features of syntactically represented images, when compared with statistical methods, provides more flexibility in some real applications, such as the extraction of deformable image patterns from complex backgrounds. In the three-dimensional object recognition context, Lin and Fu [1986] stated that it is easier to identify the visible primitive surface patches in a syntactic representation than to recognize the object directly.

In addition to that, stochastic grammars provide a probability of a given object in each class to be considered, different from other discriminant-based methods, such as decision trees, neural networks, support vector machines, or even determinist grammars, which provide only the classification outcome [Stuckelberg and Doermann 1999].

4.4. Challenges and Possible Future Directions

From the above considerations, we can outline some directions that are still open in the literature investigated.

First of all, the study of complex images remains a challenge to be overcome. Modeling, representing, segmenting, and recognizing complex objects constitute research lines that can be further explored. One of the main categories of these objects regard medical images. In the last years several computer-aided diagnosis systems have been proposed in the literature, using medical images as the main input data. Some examples are Nunes et al. [2007] and Doi [2007]. Although some of these systems are already used in certain clinical routines, the use of grammars in this context is underexplored. We believe that grammars could provide a flexible way to assist in the identification and classification of diseases, as well as to help differentiate normal and abnormal cases. Some initiatives were included in this review [Ogiela et al. 2008; Hamdi et al. 2012], but there are still many questions to be overcome. The complex nature of medical images requires in-depth comprehension of the principles of medical image modality, the disease target, and the processing goals, and on account of this, they constitute a rich research field.

Another subject to be investigated regards the development of fast retrieval of the syntactic structures. In the content-based image retrieval area, fast indexation and retrieval methods are investigated in order to decrease the time spent for searching images [Böhm et al. 2001; Petrakis et al. 2002]. Similarly, more efficient structures for indexing and storing syntactic representations can be investigated, thus contributing to the effective use of grammars for image storing and retrieval.

The combination of grammatical methods with other approaches was productive in some reported papers. As cited before, stochastic context-free grammars combined with Markov random fields result in flexible spatial characterization [Reddy et al. 2009; Zhu et al. 2009]. Fuzzy grammars and wavelet transforms were used in texture recognition, where the grammar terminal symbols are feature vectors extracted from the images after the application of wavelet transforms [Ferreira et al. 2007a, 2007b, 2009]. Similar strategies could be experimented using spatial random trees, since the framework admits feature vectors as terminals [Siskind et al. 2007]. Other combinations using different methods could be explored in order to offset their weak points.

Learning grammar rules could be further explored to automatically create the complete image models from training samples. Grammatical learning is an active research area, since there are still several open problems and computationally complex issues [Sakakibara 2005]. Learning the grammars in the higher levels of the Chomsky hierarchy presents more challenges than the grammars in lower levels. For instance, while there are known algorithms for learning finite automata from only positive or from positive and negative samples [Angluin 1992; Ron et al. 1995], the whole class of context-free grammars are not identifiable using only positive samples [Sakakibara 1995]. Some approaches for learning this class of grammars include learning context-free subclasses [Mäkinen 1992], learning from structured samples [Sakakibara 1992], and the use of heuristics [Sakakibara and Muramatsu 2000]. These and similar strategies could be proposed and applied for computer vision applications. Sainz and Sanfeliu [1996], for instance, proposed a method for learning augmented regular expressions that are able to recognize a proper subset of context-sensitive languages.

As mentioned before, all applications of L-systems surveyed here address the object construction problem. Since L-systems are parallel grammars, parsing is an issue to be addressed [Fernau 2003]. Therefore, advances in L-system parsing are needed to apply this model to object recognition problems.

Finally, there is the matter of the small number of studies in the three-dimensional domain. In fact, working in this context includes some additional difficulties, such as occlusion problems and the need to understand the perspectives of the objects, among others. However, technology advances in the last years have intensified an increase of three-dimensional interactive systems using computer graphics and virtual reality technologies in many application fields. Thus, this is a topic that deserves to be explored much more. Prusinkiewicz et al. [1988] indicated a dozen directions related to L-systems using modeling graphic objects, such as the addition of textures, the modeling of complex surfaces, and the analysis of simulation complexity. We think that in addition to these topics, interaction and its influence on the representation scenario can represent a research line that deserves further investigation.

5. CONCLUSIONS

The studies reviewed here allowed a comprehensive overview regarding the use of syntactic methods in computer vision. As can be seen, this is a very current research area, since most of the studies found were published in the last decade, especially in the last 5 years. Moreover, it can be noted that it is a very promising research line, since many of the articles studied showed a recognition rate higher than 90%.

Most of the studies conducted were in the area of recognition and creation of objects. Few studies have addressed the problem of texture recognition, changing scale, and segmentation, indicating that such problems are potential research targets. It was also noted that very few papers (19%) reported on how to learn or extract the structure of the grammar from data, which can be considered as a gap in this research area. Learning or extracting the grammar structure from data can be very useful when there is not sufficient knowledge on a certain class of images to allow one to manually design a grammar. Instead, a training image sample is used in the learning process of grammatical rules in order to recognize the patterns common to these images, which can allow characterizing them as a class.

REFERENCES

ANGLUIN, D. 1992. Computational learning theory: Survey and selected bibliography. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC'92)*. ACM, New York, 351–369. DOI: http://dx.doi.org/10.1145/129712.129746.

Böhm, C., Berchtold, S., and Keim, D. A. 2001. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Comput. Surv. 33*, 3, 322–373. DOI: http://dx.doi.org/10.1145/502807.502809.

Chanda, G. and Dellaert, F. 2004. *Grammatical Methods in Computer Vision: An Overview*. Tech. rep. Georgia Tech Institute of Technology.

Chen, H., Xu, Z. J., Liu, Z. Q., and Zhu, S. C. 2006. Composite templates for cloth modeling and sketching. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1 (CVPR'06)*. IEEE Computer Society, Washington, DC, 943–950. DOI: http://dx.doi.org/10.1109/CVPR.2006.81.

Christensen, H. I., Matas, J., and Kittler, J. 1996. Using grammars for scene interpretation. In *Proceedings of the International Conference on Image Processing*, vol. 1. 793–796 DOI: http://dx.doi.org/10.1109/ICIP.1996.561024.

Doi, K. 2007. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput. Med. Imaging Graphics 31*, 4–5, 191–211.

Fernau, H. 2003. Parallel grammars: A phenomenology. *Grammars 6,* 1, 25–87. DOI: http://dx.doi.org/10.1023/A:1024087118762.

Ferreira, M. J., Santos, C. P., and Monteiro, J. 2007a. Texture cue based tracking system using wavelet transform and a fuzzy grammar. In *Proceedings of the 5th IEEE International Conference on Industrial Informatics*, vol. 1, 393–398. DOI: http://dx.doi.org/10.1109/INDIN.2007.4384789.

Ferreira, M. J., Santos, C. P., and Monteiro, J. 2007b. Texture segmentation based on fuzzy grammar for cork parquet quality control. In *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE'07)*. 1832–1837. DOI: http://dx.doi.org/10.1109/ISIE.2007.4374884.

Ferreira, M., Santos, C., and Monteiro, J. 2009. Cork parquet quality control vision system based on texture segmentation and fuzzy grammar. *IEEE Trans. Ind. Electron. 56*, 3, 756–765.

Gao, J., Ding, X., and Zheng, J. 2000. Image pattern recognition based on examples—A combined statistical and structural-syntactic approach. In *Advances in Pattern Recognition*, F. J. Ferri, J. M. Inesta, A. Amin, and P. Pudil, Eds., Lecture Notes in Computer Science, vol. 1876. Springer, Berlin, 57–66. DOI: http://dx.doi.org/10.1007/3-540-44522-6_6.

Gidas, B. and Zelic, A. 1997. Object recognition via hierarchical syntactic models. In *Proceedingsof the 13th International Conference on Digital Signal Processing (DSP'97)*, vol. 1, 315–318. DOI: http://dx.doi.org/10.1109/ICDSP.1997.628082.

Glomb, P. 2007. Image language terminal symbols from feature analysis. In *Proceedings of the IEEE International Workshop on Imaging Systems and Techniques (IST'07)*. 1–6. DOI: http://dx.doi.org/10.1109/IST.2007.379599.

Hamdi, S., Abdallah, A. B., and Bedoui, M. H. 2012. Grammar-based image segmentation and automatic area estimation. In *Proceedings of the 16th IEEE Mediterranean Electrotechnical Conference (MELECON'12)*. 356–359. DOI: http://dx.doi.org/10.1109/MELCON.2012.6196448.

Han, F. and Zhu, S.-C. 2005. Bottom-up/top-down image parsing by attribute graph grammar. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. Beijing, China, 1778–1785, vol. 2. 1550–5499. DOI: http://dx.doi.org/10.1109/ICCV.2005.50.

Han, F. and Zhu, S.-C. 2009. Bottom-up/top-down image parsing with attribute grammar. *IEEE Trans. Pattern Anal. Mach. Intell. 31*, 1, 59–73. DOI: http://dx.doi.org/10.1109/TPAMI.2008.65.

Hemberg, M. and OReilly, U.-M. 2004. Extending grammatical evolution to evolve digital surfaces with Genr8. In *Genetic Programming*, M. Keijzer, U.-M. OReilly, S. Lucas, E. Costa, and T. Soule, Eds., Lecture Notes in Computer Science, vol. 3003. Springer, Berlin, 299–308. DOI: http://dx.doi.org/10.1007/978-3-540-24650-3_28.

Hingway, S. P. and Bhurchandi, K. M. 2011. A simple graph theoretic approach for object recognition. In *Proceedings of the 4th International Conference on Emerging Trends in Engineering and Technology (ICETET'11)*. 200–205. DOI: http://dx.doi.org/10.1109/ICETET.2011.62.

Jin, Y. and Geman, S. 2006. Context and hierarchy in a probabilistic image model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2145–2152. DOI: http://dx.doi.org/10.1109/CVPR.2006.86.

Kanungo, T. and Mao, S. 2003. Stochastic language models for style-directed layout analysis of document images. *IEEE Trans. Image Process. 12*, 5, 583–596. DOI: http://dx.doi.org/10.1109/TIP.2003.811487.

Kong, J., Barkol, O., Bergman, R., Pnueli, A., Schein, S., Zhang, K., and Zhao, C. 2012. Web interface interpretation using graph grammars. *IEEE Trans. Syst. Man Cybern., Part C Appl. Rev. 42*, 4, 590–602. DOI: http://dx.doi.org/10.1109/TSMCC.2011.2171335.

Lin, W.-C. and Fu, K.-S. 1986. A syntactic approach to three-dimensional object recognition. *IEEE Trans. Syst. Man Cybern. 16*, 405–422.

Luo, P., He, J., Lin, L., and Chao, H. 2009. Hierarchical 3D perception from a single image. In *Proceedings of the 16th IEEE International Conference on Image processing (ICIP'09)*. IEEE, Los Alamitos, CA, 4209–4212.

Mäkinen, E. 1992. On the structural grammatical inference problem for some classes of context-free grammars. *Inf. Process. Lett. 42*, 1, 1–5.

Mao, S., Rosenfeld, A., and Kanungo, T. 2003. Stochastic attributed K-d tree modeling of technical paper title pages. In *Proceedings of the 2003 International Conference on Image Processing (ICIP'03)*. 533–536. DOI: http://dx.doi.org/10.1109/ICIP.2003.1247016.

Mas, J., Jorge, J. A., Sanchez, G., and Llados, J. 2008. Representing and parsing sketched symbols using adjacency grammars and a grid-directed parser. In *Graphics Recognition. Recent Advances and New Opportunities*, W. Liu, J. Llados, and J.-M. Ogier, Eds., Lecture Notes in Computer Science, vol. 5046. Springer, Berlin, 169–180. DOI: http://dx.doi.org/10.1007/978-3-540-88188-9_17.

Mas, J., Sanchez, G., and Llados, J. 2005. An adjacency grammar to recognize symbols and gestures in a digital pen framework. In *Pattern Recognition and Image Analysis*, J. S. Marques, N. P. de la Blanca, and P. Pina, Eds., Lecture Notes in Computer Science, vol. 3523. Springer, Berlin, 115–122. DOI: http://dx.doi.org/10.1007/11492542_15.

Nunes, F. L. S., Schiabel, H., and Goes, C. E. 2007. Contrast enhancement in dense breast images to aid clustered microcalcifications detection. *J. Digital Imaging 20*, 1, 53–66.

Ogiela, L., Ogiela, M. R., and Tadeusiewicz, R. 2009. Mathematical linguistics in cognitive medical image interpretation systems. *J. Math. Imaging Vision 34,* 3, 328–340. DOI: http://dx.doi.org/10.1007/s10851-009-0151-4.

Ogiela, L., Tadeusiewicz, R., and Ogiela, M. R. 2008. Cognitive modeling in medical pattern semantic understanding. In *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering (MUE'08)*. 15–18. DOI: http://dx.doi.org/10.1109/MUE.2008.47.

Parag, T., Bahlmann, C., Shet, V., and Singh, M. 2012. A grammar for hierarchical object descriptions in logic programs. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'12)*. 33–38. DOI: http://dx.doi.org/10.1109/CVPRW.2012.6239171.

Peng, S., Liu, L., Yang, X., and Sang, N. 2008. A database schema for large scale annotated image dataset. In *Proceedings of the Congress on Image and Signal Processing (CISP'08)*. 57–62. DOI: http://dx.doi.org/10.1109/CISP.2008.49.

Petrakis, E. G. M., Faloutsos, C., and Lin, K.-I. 2002. ImageMap: An image indexing method based on spatial similarity. *IEEE Trans. Knowl. Data Eng. 14*, 5, 979–987. DOI: http://dx.doi.org/10.1109/TKDE.2002.1033768.

Prusinkiewicz, P., Lindenmayer, A., and Hanan, J. 1988. Development models of herbaceous plants for computer imagery purposes. *SIGGRAPH Comput. Graph. 22*, 4, 141–150. DOI: http://dx.doi.org/10.1145/378456.378503.

Qu, H., Zhu, Q., Zeng, L., Guo, M., and Lu, Z. 2008. Automata-based L-Grammar extraction from multiple images for virtual plants. In *Proceedings of the 3rd International Conference on Bio-Inspired Computing: Theories and Applications (BICTA'08)*. 89–96. DOI: http://dx.doi.org/10.1109/BICTA.2008.4656709.

Reddy, H. T., Karibasappa, K., and Damodaram, A. 2009. Probabilistic parser for face detection. In *Proceeding of International Conference on Methods and Models in Computer Science (ICM2CS'09)*. 1–7. DOI: http://dx.doi.org/10.1109/ICM2CS.2009.5397986.

Ron, D., Singer, Y., and Tishby, N. 1995. On the learnability and usage of acyclic probabilistic finite automata. In *Proceedings of the 8th Annual Conference on Computational Learning Theory (COLT'95)*. ACM, New York, 31–40. DOI: http://dx.doi.org/10.1145/225298.225302.

Rothrock, B. and Zhu, S.-C. 2011. Human parsing using stochastic and-or grammars and rich appearances. In *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops'11)*. 640–647. DOI: http://dx.doi.org/10.1109/ICCVW.2011.6130303.

Sainz, M. and Sanfeliu, A. 1996. Learning bidimensional context-dependent models using a context-sensitive language. In *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 4, 565–569. DOI: http://dx.doi.org/10.1109/ICPR.1996.547628.

Sakakibara, Y. 1992. Efficient learning of context-free grammars from positive structural examples. *Inf. Comput. 97*, 1, 23–60. DOI: http://dx.doi.org/10.1016/0890-5401(92)90003-X.

Sakakibara, Y. 1995. Grammatical inference: An old and new paradigm. In *Algorithmic Learning Theory*, K. P. Jantke, T. Shinohara, and T. Zeugmann, Eds., Lecture Notes in Computer Science, vol. 997. Springer, Berlin, 1–24. DOI: http://dx.doi.org/10.1007/3-540-60454-5_25.

Sakakibara, Y. 2005. Learning context-free grammars using tabular representations. *Pattern Recogn. 38*, 9, 1372–1383. DOI: http://dx.doi.org/10.1016/j.patcog.2004.03.021.

Sakakibara, Y. and Muramatsu, H. 2000. Learning context-free grammars from partially structured examples. In *Grammatical Inference: Algorithms and Applications*, A. L. Oliveira, Ed., Lecture Notes in Computer Science, vol. 1891. Springer, Berlin, 229–240. DOI: http://dx.doi.org/10.1007/978-3-540-45257-7_19.

Schlecht, J., Barnard, K., Spriggs, E., and Pryor, B. 2007. Inferring grammar-based structure models from 3D microscopy data. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'07)*. 1–8. DOI: http://dx.doi.org/10.1109/CVPR.2007.383031.

Shet, V., Singh, M., Bahlmann, C., and Ramesh, V. 2009. Predicate logic based image grammars for complex pattern recognition. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops'09)*. 7. DOI: http://dx.doi.org/10.1109/CVPRW.2009.5204328.

Shilman, M., Liang, P., and Viola, P. 2005. Learning nongenerative grammatical models for document analysis. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'05)*, vol. 2, 962–969. DOI: http://dx.doi.org/10.1109/ICCV.2005.140.

Si, Z. and Zhu, S.-C. 2011. Unsupervised learning of stochastic AND-OR templates for object modeling. In *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops'11)*. 648–655. DOI: http://dx.doi.org/10.1109/ICCVW.2011.6130304.

Siddiqi, K. and Kimia, B. B. 1996. A shock grammar for recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*. 507–513. DOI: http://dx.doi.org/10.1109/CVPR.1996.517119.

Siddiqi, K., Shokoufandeh, A., Dickenson, S. J., and Zucker, S. W. 1998. Shock graphs and shape matching. In *Proceedings of the 6th International Conference on Computer Vision*. 222–229. DOI: http://dx.doi.org/10.1109/ICCV.1998.710722.

Sipser, M. 2006. *Introduction to the Theory of Computation* (2nd ed.). Thomson Course Technology.

Siskind, J., Sherman Jr, J., Polak, T., Harper, M., and Bouman, C. 2007. Spatial random tree grammars for modeling hierarchical structure in images with regions of arbitrary shape. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 9, 1504–1518.

Soltanpour, S. and Ebrahimnezhad, H. 2010. Learning novel object parts model for object categorization. In *Proceedings of the 5th International Symposium on Telecommunications (IST'10)*. 796–800. DOI: http://dx.doi.org/10.1109/ISTEL.2010.5734131.

Stuckelberg, M. V. and Doermann, D. 1999. On musical score recognition using probabilistic reasoning. In *Proceedings of the 5th International Conference on Document Analysis and Recognition (ICDAR'99)*. 115–118. DOI: http://dx.doi.org/10.1109/ICDAR.1999.791738.

Subramanian, K. G., Mary, A. R. S., and Dersanambika, K. S. 2005. Splicing array grammar systems. In *Proceedings of the International Conference on Theoretical Aspects of Computing (ICTAC'05)*, D. Hung and M. Wirsing, Eds., Lecture Notes in Computer Science, vol. 3722. Springer, Berlin, 125–135. DOI: http://dx.doi.org/10.1007/11560647_8.

Sun, R., Jia, J., Li, H., and Jaeger, M. 2009. Image-based lightweight tree modeling. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI'09)*. ACM, New York, 17–22. DOI: http://dx.doi.org/10.1145/1670252.1670258.

Takada, Y. 1988. Grammatical inference for even linear languages based on control sets. *Inform. Process. Lett. 28*, 4, 193–199. DOI: http://dx.doi.org/10.1016/0020-0190(88)90208-6.

Toshev, A., Mordohai, P., and Taskar, B. 2010. Detecting and parsing architecture at city scale from range data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*. 398–405. DOI: http://dx.doi.org/10.1109/CVPR.2010.5540187.

Trzupek, M., Ogiela, M. R., and Tadeusiewicz, R. 2009. Image content analysis for cardiac 3D visualizations. In *Knowledge-Based and Intelligent Information and Engineering Systems*, J. D. Velásquez, S. A. Ríos, R. J. Howlett, and L. C. Jain, Eds., Lecture Notes in Computer Science, vol. 5711. Springer, Berlin 192–199. DOI: http://dx.doi.org/10.1007/978-3-642-04595-0_24.

Trzupek, M., Ogiela, M. R., and Tadeusiewicz, R. 2011. Intelligent image content description and analysis for 3D visualizations of coronary vessels. In *Intelligent Information and Database Systems*, N. T. Nguyen, C.-G. Kim, and A. Janiak, Eds., Lecture Notes in Computer Science, vol. 6592. Springer, Berlin, 193–202. DOI: http://dx.doi.org/10.1007/978-3-642-20042-7_20.

Tylecek, R. and Sara, R. 2011. Modeling symmetries for stochastic structural recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops'11)*. 632–639. DOI: http://dx.doi.org/10.1109/ICCVW.2011.6130302.

Wang, Q. and Jiang, Z. 2009. A grammatical framework for building rooftop extraction. In *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing (IGARSS'09)*. III–334–III–337. DOI: http://dx.doi.org/10.1109/IGARSS.2009.5417768.

WANG, W., POLLAK, I., BOUMAN, C. A., AND HARPER, M. P. 2005b. Classification of images using spatial random trees. In *Proceedings of the 2005 IEEE/SP 13th Workshop on Statistical Signal Processing*. 449–452. DOI: http://dx.doi.org/10.1109/SSP.2005.1628637.

WANG, W., POLLAK, I., WONG, T.-S., BOUMAN, C. A., HARPER, M. P., AND SISKIND, J. M. 2006. Hierarchical stochastic image grammars for classification and segmentation. *IEEE Trans. Image Process. 15*, 10, 3033–3052. DOI: http://dx.doi.org/10.1109/TIP.2006.877496.

WANG, Y., BAHRAMI, S., AND ZHU, S.-C. 2005a. Perceptual scale space and its applications. In *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05)*. Vol. 1, 58–65. DOI: http://dx.doi.org/10.1109/ICCV.2005.187.

WU, Y. AND BIAN, H. 2009. Image segmentation integrating generative and discriminative methods. In *Proceedings of the International Conference on Web Information Systems and Mining (WISM'09)*. 769–774. DOI: http://dx.doi.org/10.1109/WISM.2009.159.

YAO, B., YANG, X., AND WU, T. 2009. Image parsing with stochastic grammar: The Lotus Hill dataset and inference scheme. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops'09)*. DOI: http://dx.doi.org/10.1109/CVPRW.2009.5204331.

ZABOLI, H. AND RAHMATI, M. 2007. An improved shock graph approach for shape recognition and retrieval. In *Proceedings of the 1st Asia International Conference on Modelling Simulation (AMS'07)*. 438–443. DOI: http://dx.doi.org/10.1109/AMS.2007.13.

ZHU, L., CHEN, Y., AND YUILLE, A. 2009. Unsupervised learning of probabilistic grammar-Markov models for object categories. *IEEE Trans. Pattern Anal. Mach. Intell. 31*, 1, 114–128. DOI: http://dx.doi.org/10.1109/TPAMI.2008.67.

ZHU, S.-C. AND MUMFORD, D. 2006. A stochastic grammar of images. *Found. Trends. Comput. Graph. Vis. 2*, 4, 259–362. DOI: http://dx.doi.org/10.1561/0600000018.

Received March 2012; revised December 2012; accepted May 2013