Attribute Multiset Grammars for Global Explanations of Activities

Dima Damen
http://www.comp.leeds.ac.uk/dima
David Hogg
http://www.comp.leeds.ac.uk/dch
School of Computing,
University of Leeds

Most activity recognition techniques focus on recognizing a single activity given a video sequence. Realistic surveillance involves multiple interleaved activities, often extending over a long temporal duration. In these situations, the activities are often mutually constrained. For example, a person entering a building can be observed departing only once at a later time. In visual interpretation, these constraints can be exploited to disambiguate uncertain visual data through seeking a globally consistent explanation. However, a general way to formalise the set of globally consistent explanations for a given domain is not yet available.

We use Attribute Multiset Grammars (AMG) as a formal representation for a domain's knowledge to encode intra- and inter-activity constraints. Each rule in an AMG rewrites a nonterminal symbol as a multiset. It combines this with attributes to extract meaning from parse trees and to constrain the application of rules. A parse tree for a set of detections, according to AMG, provides a feasible global explanation.

AMGs were introduced in [3] for representing the constituents and layout of a picture. We define an AMG G = (N, T, S, A, P) where N is the set of nonterminal symbols representing activities and compound events, T is the set of terminal symbols representing primitive events, S is the start symbol (S $\in$ N), A(x) is a set of attributes defined for the symbol $x \in N \cup T$. $A(x) = A_0(x) \cup A_1(x)$, where $A_0(x)$ is the set of *synthetic* attributes which have calculated values for all primitive events, and $A_1(x)$ is the set of *inherited* attributes which are explanation-related like the number of activities in which the event participates or a textual description. Finally, P is the set of production rules. Each production rule decomposes a compound event into simpler events, and is associated with attribute rules and constraints. We distinguish between two types of constraints; synthetic and inherited constraints. Synthetic constraints define intra-activity constraints. Inherited constraints, on the other hand, govern relationships between activities, such as sharing events between activities. For example, to recognize the event of picking a person up by a car, the person can be picked up once, while the car can still pick up other people. Figure 1 shows an abstract AMG, while Figure 2 presents a parse tree given a set of detections. The parse tree represents a global feasible explanation.

Nonterimanls (N):	S	Start symbol	
	A	compound event 1	
	B	compound event 2	
Terminals (T):	α	primitive event 1	
	β	primitive event 2	
	γ	primitive event 3	
Attributes (A):			
attribute name	**type**	**domain**	**defined for**
time	$\in A_0$	int	$\{\alpha, \beta, \gamma$, A, B$\}$
count	$\in A_1$	int	$\{\beta$, B$\}$

Production Rules (P):

	Rule (r)	Attribute Rules (M)			Attribute Constraints (C)		
p_1	S $\rightarrow$ A*, B*, $\alpha^\star, \gamma^\star$						
p_2	A $\rightarrow$ α, B	A.time	=	α.time+B.time	α.time	$<$	B.time
		B.count	=	1	B.count	$\neq$	1
p_3	B $\rightarrow$ β, γ	B.time	=	γ.time	β.time	$<$	γ.time
		β.count	=	1	β.count	$\neq$	1

Figure 1: Example of an Attribute Multiset Grammar

To find the best parse tree given the detections, we build a Bayesian network (BN) as explained in our previous work [2], with conditional links between events and their associated observations, between compound events and their constituent events, and between nodes and a deterministic random variable when enforcing consistency. The desired explanation is the MAP for this network. Figure 2 shows a labeled Bayesian network that corresponds to the parse tree. Notice that a hidden random variable (RV) represents each possible nonterminal in a parse tree, and is labeled true if the nonterminal appears in the parse tree, and false otherwise. The paper includes an algorithm for building this BN out of a set of detections and an AMG.

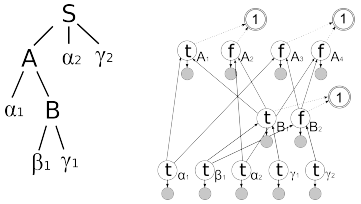

Figure 2: A parse tree and the corresponding labeling of the BN

The paper presents AMGs for two surveillance tasks. The first is the Bicycles problem, previously presented informally in [2]. The task is to correctly associate people to the bicycle they have dropped or picked, and to link picks to earlier drops. The second is the challenging problem of associating pedestrians and carried objects entering and departing a building. For example, we wish to recognize an individual entering a building with a bag and departing without it. Figure 3 shows a parse tree of the grammar and the corresponding labeled Bayesian Network.

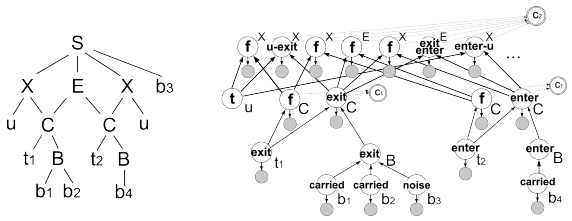

Figure 3: A sample parse tree and labelled BN for the *Enter-Exit* problem

The technique was tested on a full day (12 hours) outside a building entrance. 326 people were tracked around the entrance after manually rejecting groups of people walking together. The baggage detector from [1] was run on the dataset resulting in 429 candidate bags. The paper details the attributes selected for the task. The results show that global explanations outperform local analysis, using the same information. Three search techniques were compared: Greedy global explanation, Multiple Hypothesis Tree, and using Reversible Jump Markov Chain Monte Carlo [2]. As for the Bicycles problem [2], RJMCMC gave the best results (Figure 4). The challenge in future is to investigate the generality of the AMG approach to similar problems.

Figure 4: Correctly paired sequences when global constrained explanations are considered

[1] Dima Damen and David Hogg. Detecting carried objects in short video sequences. In *European Computer Vision Conference (ECCV)*, volume 3, pages 154–167, 2008.

[2] Dima Damen and David Hogg. Recognizing linked events: Searching the space of feasible explanations. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 927–934, 2009.

[3] Eric Gollin. *A Method for the Specification and Parsing of Visual Languages*. PhD thesis, Brown University, 1991.