Predicate Logic based Image Grammars for Complex Pattern Recognition

Vinay Shet Maneesh Singh Claus Bahlmann Visvanathan Ramesh
Siemens Corporate Research, 755 College Rd East, Princeton, NJ
{vinay.shet; maneesh.singh; claus.bahlmann; visvanathan.ramesh}@siemens.com

Reliably detecting patterns in visual data has been the primary goal of computer vision research for several years. Such patterns could be strictly spatial in nature, like static images of pedestrians, bicycles, airplanes etc., or spatio-temporal in nature, like patterns of human or vehicular activity over time. Lately, there has been increased interest in using image grammars to represent and parse visual images and detect such patterns [5, 4]. Predicate logic based reasoning approaches provide one such means to formally specify domain knowledge, and manipulate symbolic information, to explicitly reason about the presence of different patterns of interest. Such logic programs provide a language to model hierarchical, compositional patterns, and combine contextual information with the detection of low level parts, via conjunctions, disjunctions and negations.

While formal logical reasoning approaches have long been used in automated theorem proving, constraint satisfaction, and computational artificial intelligence, historically, their use in the field of computer vision has remained limited. One of the primary inhibiting factors to a successful integration of computer vision and first order predicate logic has been the design of an appropriate interface between the binary-valued logic and probabilistic vision output. Bilattices, algebraic structures introduced by Ginsberg [2], provide a means to design exactly such an interface to model uncertainties for logical reasoning.

Unlike traditional logics, predicate logics extended using the bilattice-based uncertainty handling formalism, associate uncertainties with both logical rules and observed logical facts. These uncertainties are taken from, and semantically interpreted within, a set structured as a bilattice. Modeling uncertainties in the bilattice facilitates independent representation of both positive and negative constraints about a proposition, and furthermore provides tolerance for contradictory data inherent in many real-world applications. Performing inference in such a framework is also, typically, computationally efficient. Additionally, as with any logic based approach, it is straightforward to generate proofs or justifications permitting direct analysis of the final solution in a linguistic form.

In [3], we addressed the problem of detecting partially occluded humans in static images by capturing knowledge about contextual cues, scene geometry and human body constraints in the form of predicate logic rules and applying them to the output of low level, human body part detectors. Positive and negative information from different rules, and uncertainties from detections were integrated within the bilattice framework. Figure 1 shows a sample result of detected humans and their corresponding uncertainty values.

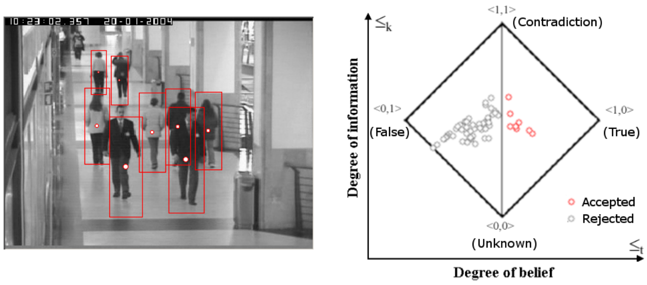

Figure 1. (Left) Sample result for human detection on the CAVIAR[1] dataset. (Right) Computed uncertainty value (for all human hypotheses in left image) plotted in the bilattice space.

In this work, we have extended the work reported in [3] to detect complex objects in aerial images. Such objects, e.g. surface-to-air missile launcher sites, are highly variable in appearance and can only be characterized by their functional design and surrounding context, such as physical arrangement of access structures. Constraints in acquiring sufficient annotated data for learning make it challenging for purely data driven approaches to adequately generalize. In this work, structure arising from functional requirements and surrounding context has been encoded using predicate logic based grammars. Observation and model uncertainties have been integrated within the bilattice framework.

In this work we have also proposed a method to automatically optimize weights associated with these logical rules. Automated logical rule weight learning is an important aspect of the application of such systems in the computer vision domain. The proposed approach casts the instantiated inference tree as a knowledge-based neural network, interprets rule uncertainties as link weights in the network, and applies a constrained, back-propagation (BP) algorithm to converge upon a set of weights for optimal performance. The BP algorithm has been accordingly modified to compute local gradients over the bilattice specific inference operation and respect constraints specific to vision applications. Both extensions have been evaluated over real and simulated data with favorable results.

References

[1] CAVIAR http://homepages.inf.ed.ac.uk/rbf/caviar/.
[2] M. L. Ginsberg. Multivalued logics: Uniform approach to inference in artificial intelligence. *Comput. Intelligence*.
[3] V. Shet, J. Neumann, V. Ramesh, and L. Davis. Bilattice-based logical reasoning for human detection. In *CVPR*, 2007.
[4] S. Todorovic and N. Ahuja. Learning subcategory relevances to category recognition,. In *CVPR*, 2008.
[5] S. Zhu and D. Mumford. A stochastic grammar of images. *Founds. and Trends in Comp. Graphics and Vision*, 2007.