

Probabilistic Ontology Induction for Generalized Video Understanding
PI: Jason J. Corso, SUNY at Buffalo
Final Report for Grant Ending June 30, 2010

This report explains the programmatic, technical and funding aspects for Dr. Corso's Phase I CSSG award. This is the final report for the grant covering March 5, 2009 through June 30, 2010.

Programmatic

Attendance/participation in sessions

Dr. Corso participated in all four CSSG panels through the year. During all sessions, Dr. Corso made every effort to actively participate in the scheduled activities by engaging briefers, other panelists, and mentors whenever possible. These activities are clearly a once in a lifetime opportunity to experience and learn about the Department of Defense (DoD) and the Intelligence Community (IC) from the most junior through most senior ranks.

In addition to required CSSG panels, Dr. Corso made three trips to interact with contacts he made through the CSSG program.

The first such trip was on July 2009 to DoD/IC offices and to the CACI offices in MD. Dr. Corso has significant interest and success in collaboration with these groups on the intelligence and forensics problems. This visit was primarily an introductory one during which Dr. Corso gave a briefing on his research capabilities and goals. The team gave Dr. Corso a set of briefings to thoroughly teach him their mission and activities.

The second trip was on August 2009 to the DC region again. This was a working session during which Dr. Corso sat with analysts and watched them work through his forensic analysis of recently acquired data sets. During this session, Dr. Corso learned about the current practices and problems facing the analysts. The need for improved forensic tools is paramount as the massive datasets, primarily images and video, are added to the corpus. Furthermore, he learned of the importance that the human video analyst plays in the forensic role.

On the third such trip, Dr. Corso travelled to three DoD/IC installations in February 2010 to further his education on defense and intelligence science; the visits took place from Feb. 16-19, 2010. During each visit, Dr. Corso delivered a seminar about his research work, his DARPA CSSG involvements, his research lab and capabilities; he engaged in research-oriented discussions with the various personnel (some of which have resulted in follow-on projects and proposals, see below); and he participated in briefings delivered by staff members at each installation designed to educate him about the critical technical challenges faced.

Expected impact of program participation on research

The four CSSG sessions have had a great impact on Dr. Corso's understanding of the problems facing the DoD/IC currently. They have impact his research at both a programmatic low-level and a broader, high-level.

At the low-level, the sessions have demonstrated the critical DoD/IC needs in Dr. Corso's research areas of computer vision and machine learning, primarily from a high-, or semantic-level. The three direct avenues his research can impact the DOD: automatic analysis of images and video for targeting and surveillance (e.g., Predator feeds), intelligence mining via semantic analysis of images and video (e.g., media exploitation and Defense forensics), adversarial behavior modeling from multiple sources of data. In the near-term, these experiences has greatly impacted Dr. Corso's plans for Defense-relevant research, including

Phase II funding. The initial plan in his Phase I research was to induce a probabilistic ontology directly from video of a particular phenomenon. The human role was left to the end to supply the semantics. However, he has realized that the human needs to be involved throughout the process. This has led him to investigate a new branch of clustering called active clustering, which involves the human throughout the clustering process. Ultimately patterns are discovered in the data, which can then be used for model building and inference, but few or no assumptions are made beforehand and the human expert is involved throughout the entire process. This is a new, unexplored, and unpublished set of ideas that is the topic of his Phase II project, which was funded and began in late June 2010.

At the high-level, they have further solidified his research direction in high-level computer vision and have helped solidify the following four driving research questions that now represent the cornerstones of his research agenda:

1. How to use principled hierarchical and probabilistic structures to model complex real-world phenomena?
2. How to handle the massive data glut for machine learning yet require little or no user labeling?
3. How to incorporate prior high-level knowledge (semantics, context, etc.) during both learning and inference?
4. How to understand and enhance the role of the user in active semi-supervised learning scenarios?

Enabled Collaborations and Opportunities

DoD/IC These research directions are allowing Dr. Corso to forge good collaborations with members of the DoD/IC. The core problems are downstream exploitation and analysis. Dr. Corso is also making collaborations with both language and vision scientists. The main contribution Dr. Corso can make is the development of a powerful forensics workbench for the analyst that utilizes not only cutting edge vision research but also ontologies.

ARL Dr. Corso has forged a collaboration with the Asset Behavior and Control Branch of the Computing Sciences Division of the Army Research Labs, led by Dr. Barbara Broome. His main collaborator there is Dr. Phil David. During Dr. Corso's visit to the ARL in February, he and Dr. David conceptualized a new research application for high-level computer vision in mobile robotics. This conceptualization has led to two funding proposals thus far, one of which was nominated for the PECASE award by the Army Research Office.

NGA Innovision Dr. Corso visited the NGA in Feb. 2010 to learn more about the NGA. His primary contact was introduced through the DARPA CSSG program. The primary driving work with the NGA is the use of Dr. Corso's latent hierarchical graphical models for capturing complex physical and human spatial processes.

CACI Dr. Corso has visited the Maryland CACI office to meet with Dr. Kristen Summers and her team twice during the year. The CACI team is primarily interested in language and document processing, which greatly complements Dr. Corso's interests in imaging and visual processing. We have jointly submitted a research proposal to IARPA in Sept. 2009 and are considering other avenues of interaction.

CIA / IC Postdoc Fellowship Dr. Corso learned about the IC Postdoc Fellowship program during the CSSG panel trip to the CIA. He submitted a proposal entitled "Semantic Video Summarization" that pulled from his CSSG experiences as well as his research expertise in ontologies for visual processing. The proposal was funded. Dr. Corso is in the process of selecting the postdoc, currently. This is a clear win for Dr. Corso from the CSSG program.

DARPA Mind's Eye Dr. Corso is the PI on a new DARPA Mind's Eye grant that also significantly pulled from his CSSG experiences and his collaborations with the ARL (that was made during the CSSG session). His CSSG involvement quite clearly positively impacted the research team and research proposal he was able to build.

Technical

In the technical section of the report, we outline our work in the context the project goals and proposed actions from the proposal.

Project Goals

The main objective of this project was "to understand how probabilistic ontologies of visual phenomena can be induced directly from video thereby revolutionizing our ability to rapidly learn a probabilistic low-to-high level domain model directly from data and use it to automatically infer a comprehensive yet parsimonious semantic description with quantitative underpinnings of video."

This objective has been partially achieved but not to the expected effect of the PI. The proposed research is based on the observation that *similar objects coarsen similarly* based on local appearance measures, and that this behavior can be used to drive the induction of probabilistic ontologies. The key here is that we need to properly define probabilistic ontology. First, an ontology is a specification of the knowledge and structure of a particular phenomena, which usually contains a description of all of the elements of the phenomena, their attributes, and the various ways in which they can interact. The ontology itself can live at a generic level (e.g., cars, trees) and at a specific level (e.g., this red sports cat, that spruce), at which time many would consider it a database. The idea of a probabilistic ontology is to incorporate the uncertainty about the elements and their relations. The validity of such probabilistic ontologies, from a philosophical point of view is an actively debated and unanswered question. We have operated under the premise that at the generic level, there is no uncertainty (about the elements and their relations, unless the ontology is specifying about uncertainty itself or some process with uncertainty, such as visual inference) but there will be uncertainty at specific level.

An ontology is hence about human semantics and the way humans understand and describe certain phenomena. This is clearly not readily attainable by an induction process without a human in the loop. Indeed, this is a primary lesson-learned of our research and is complemented by the observations made during PI's interactive working sessions with intelligence analysts (i.e., the human is invaluable to image and video analysis from an intelligence perspective). A fundamental aspect of the PI's Phase II proposal is the role of the human in an interactive clustering process.

On the other hand, a probabilistic ontology which may be tied to an ontology of generics in some way is more readily achievable by an inductive learning procedure. The induced model then more readily bridges the signal-symbol gap because the semantics themselves are those inherent directly within the data. Our approaches to adaptive coarsening and probabilistic models on those hierarchies have helped us to achieve this goal. The methods are directly usable in many defense relevant problems, such as visual entity extraction. We do note, however, that we still need some notion of annotation either interactively by the human or proactively by the human on training data in order to make the final link from the probabilistic models to the human semantics.

Technical Performance

Adaptive Hierarchical Graphs for Modeling

Action A1. We will extend our existing adaptive coarsening method from images to videos.

This action was completed by the PI and RA Albert C during the summer quarter 2009. C extended Corso's past work on adaptive coarsening of images to videos. He considered the video as a spatiotemporal cube and derived similarity measures that operated across both space and time. The hierarchies he developed were primarily based on algebraic multigrid approaches and are hence represented as multilevel graph. C also developed a general and powerful C++ library of these adaptive coarsening routines. The methods will continue to be used in the PI's defense relevant research.

Action A2. We will define a set of discriminative features to describe each hierarchical region.

This action was completed by the PI and RA Albert C also during the summer quarter 2009. The measures of similarity across space and time were the primary activity to specify the set of discriminative features. We also developed various other appearance-based mechanisms, such as sparse-coding (which are used in Action A3) and texture, shape and hierarchical descriptors.

Probabilistic Ontology Induction

Action A3. We will develop a graph-clustering algorithm to isolate stable repeating structures in the hierarchies.

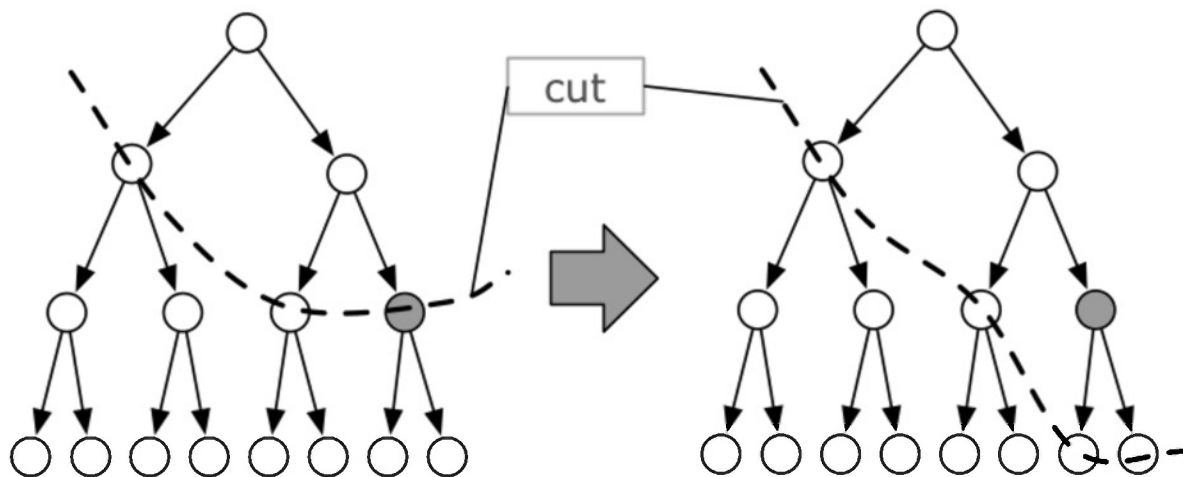
The PI and RA Caiming X spent a significant amount of the active project time on this action as it proved to present the most significant technical challenge of the work. The key question is how nodes in the spatiotemporal hierarchies should be compared. We have emphasized a sparse-coding approach to establish a basis for the regions and then make a similarity comparison based on the sparse-coded basis. This approach seemed to prove sufficient for the current work, but we did not have sufficient time to make a thorough experimental analysis of it.

Action A4. We will design a theoretically sound approach for how to query an oracle (human or computer method) during induction to identify new elements (of the ontology).

The PI and RA Caiming X investigated this action and developed a plausible solution to it, which is briefly described below. However, during the course of our work, we realized the magnitude of this problem (with A3). After our intense investigation, we based the majority of the Phase II proposal on it and feel a rigorous, principled solution to this action is years ahead, but once achieved will be a significant fundamental contribution to machine intelligence.

The basic approach we considered for this action was one of defining a cut through the graph hierarchy. Recall the graph hierarchy encapsulates a multiscale decomposition of the image, and, when performed in feature space disregarding spatial continuity, the multiscale decomposition of the image represents a set of possible clusters. We have assumed, for now, that the graph is a tree. The cut through the graph is defined as the set of nodes currently in our cluster set. In other words, the cut defines the current elements from the image or video that are to be elements in our ontology. The human is asked questions, based on the current cut, of the form: Should these regions in the image be grouped into the same cluster? The user is given the image with various parts (based on the hierarchy and the selected node) highlighted. The user can then answer yes, no, or not sure and the cut will update accordingly. A representative figure is shown below where the cut is drawn in a dotted line and the node in question is colored gray. Here, the user answers 'no' and the cut moves to its children. The node used in the query is selected based on its relative entropy with respect to the rest of the tree. We have experimented with different objective functions to drive which node to next query on that are based on node purity (a term borrowed from the decision tree literature) defined

using a node's sub-tree appearance distribution and its entropy, cross-node similarity measures. At this time, we have not arrived at a final definition for the objective criterion.



Action A5. We will build a modeling framework that includes all elements and induced interrelations in a mathematically coherent manner.

The PI developed a modeling framework for these adaptive multilevel hierarchies and semantically grounded probabilistic methods that is defined as a layered set of random fields (whose nodes do not reside on a lattice). We have attached an unpublished working paper entitled "Random Multilevel Fields for Image Labeling" based on the outcome of this Action.

Dynamic Graph-Shifts Inference

Action A6. We will extend our graph-shifts algorithm to utilize the induced probabilistic ontologies to drive inference.

The PI and RA Albert C have extended the graph-shifts algorithm to be applied to the adaptively coarsened video hierarchies, which are instances of the probabilistic ontologies. We have attached an unpublished working paper entitled "Video Graph-Shifts" based on the outcome of this Action.

Funding

Here, the dollar values are listed only.

Incurred expenses: \$99,658.63.

Incurred expenses as % of obligated funding: $\$99,658.63/\$99,670.00 = 99.98\%$.

Invoiced expenses: \$99,658.63.