

# RESEARCH STATEMENT

Raffay Hamid (raffay@cc.gatech.edu)

I want to build systems that can improve our everyday lives by being able to perceive, learn, and predict what is happening around them. My research goals are driven by the tremendous synergy between the ubiquity of sensors and the computational power of machines. I believe the missing link to fully capitalize this potential is the set of computational mechanisms that would help interpret the otherwise nonsensical sensory data. These mechanisms would play a vital role in building smarter robots, monitoring people's health as they age, and preventing crime through improved surveillance. Over the last few years I have been exploring such mechanisms, particularly for the perceptual understanding of everyday human activities.

## Background and Current Work

Human activities can be interpreted at different levels of abstraction. Consider for instance the activity of a person making an omelet in a household kitchen. Captured through a video camera, one way of looking at this activity may be using the motion of various color-blobs in the scene, detected directly through the raw pixel-data. Another way to look at it may be in terms of certain semantically meaningful activity-descriptors, such as beating some eggs, heating oil, frying the eggs, *etc.* The granularity at which one chooses to decompose an activity offers a tradeoff between the expressiveness of its descriptors, and their perceptual detectability. This tradeoff is a function of activity dynamics, and the environment in which it is being performed. I am particularly interested in computational frameworks which facilitate the analysis of human activities that have complex dynamics, and are performed in everyday environments<sup>1</sup>.

## A Data-Driven Approach Towards Activity Modeling

A key step towards this end is finding appropriate models for activity representation. Traditional approaches for activity modeling assume prior knowledge about the structure of activities, based on which explicitly defined models are learned in a supervised manner. However, such activity structure is generally not completely known *a priori*. It is therefore imperative to find representations that facilitate learning of this structure with minimal supervision.

I approach this challenge with the intuition that the world around us poses a certain set of spatial and temporal constraints on the way we generally execute our activities. For instance, one has to open a fridge before one can get milk out of it. Similarly, one must turn a stove on before one can use it to fry eggs, *etc.* My hypothesis is that these constraints can be used to define a certain set of perceptually detectable activity-descriptors. Moreover, I believe that only a few subsequences of such descriptors are sufficient to uniquely encode the structure of different activities. This idea of representing a sequence as a conjunction of its sequential features has previously been used to capture the structure of various temporal processes, including speech, text documents, and protein sequences.

In this regard, I have focused on representations that model activities in terms of their fixed and variable-length sequential features [1] [3]. I have used these representations to model activities in environments such as a residential house, a loading dock area and a household kitchen. My research findings demonstrate the importance of encoding activity structure at a select set of temporal scales. They also highlight the tradeoff between the rigidity with which a representation encodes the activity structure, and its sensitivity to sensor noise. I believe that such data-driven approach towards activity representation can help discover and characterize human activities, and learn typical behaviors critical for detecting irregular occurrences in an environment.

## Unsupervised Discovery of Human Behaviors

Categories and concepts are crucial for making sense of the world around us. In a household kitchen for instance, concepts such as washing dishes, frying an omelet, or having some cereal *etc.*, are important to analyze previously unobserved kitchen activities. While recent increase in the availability of sensors has resulted in vast amounts of readily available activity-data, most of this data is unlabeled, and therefore cannot be used to learn models of activity-classes in a supervised framework.

My approach to this challenge stems from the feature-based view of concepts and categories. This view posits that members of an activity-class generally share a set of common properties that make them perceptually similar to each other. For instance, activities of frying omelets look similar to each other as they mostly require events such as beating eggs followed by frying them. I believe that my representation of modeling activities as conjunctions of their sequential features supports a notion of their perceptual similarity that can be used for the unsupervised discovery [1] and characterization [2] of various types of human behaviors.

---

<sup>1</sup>Related material for this document can be found online at: [www.cc.gatech.edu/~raffay/research.htm](http://www.cc.gatech.edu/~raffay/research.htm)

Using graph-partitioning techniques I have tried to corroborate this hypothesis for a variety of everyday environments. User-studies of the discovered human behaviors in these environments suggest that most of the discovered activity-classes are semantically meaningful to human observers, indicating successful manipulation of the bias induced in the system in terms of the activity-descriptors.

### Anomalous Activity Detection

Finding behaviors that are in some sense anomalous is crucial for the purposes of monitoring and surveillance. Since anomalies are rare occurrences with large variation amongst them, traditional approaches that attempt to learn explicitly defined models of anomalies do not generalize well.

I approach the problem of finding anomalous behaviors from a detection rather than a recognition based perspective. As the notion of anomaly is closely related to what is meant by regular, I have modeled anomalies as activities that deviate from behaviors perceived as regular in an environment. Using the discovered activity classes to learn the concept of regularity in an environment, I have tried to detect anomalies that deviate from regular behaviors [1].

My research findings demonstrate the importance of formalizing differences of anomalous activities from regular behaviors at multiple temporal scales [3]. Moreover, since all deviations from regular behaviors are not necessarily interesting, I believe that taking a purely detection-based perspective towards finding anomalies can in fact be too general. One way of striking a balance between the brittleness of an exclusively recognition-based perspective and the generality of a purely detection-based view towards anomalies, is to learn certain domain specific constraints on what makes various irregularities truly alarming. How these constraints should be modeled and learned for different environments remains an open question.

### Perceptual Understanding of Human Behaviors – A Research Agenda

I am interested in computational mechanisms that would enable systems to be perceptually aware of what is happening around them. I believe the competence of these mechanisms fundamentally depends on the representation and manipulation of the various types of perceptual information about our environment.

I view this challenge of finding the right perceptual bias from a learning-based perspective. Unlike traditional knowledge-based approaches, I am interested in minimally supervised mechanisms that would allow a system to use its sensory data in order to learn characterizations that best inform its inference. Motivated by the theory of *embodied cognition*, such a data-driven approach of learning perceptual bias focuses more on the detection and learnability of concept-characterizations, rather than their human interpretability. It therefore facilitates the acquisition of detectable, robust, and adaptable characterizations that can be used to learn concepts of increasing complexity.

So far, I have focused mostly on systems where some basic perceptual bias is encoded by an expert only at the start of the learning process. In the future, I want to focus more [4] on designing mechanisms in which this bias can be interactively added to the system to further improve its inference. Such mechanisms could for instance incorporate an expert's assistance to dynamically cluster new behaviors in an environment, modify an existing notion of anomalous behaviors, and model certain attributes of an environment that may change over time.

I am particularly interested in using this active-learning paradigm for systems that can analyze group behaviors. These systems would for instance allow teams of fire-fighters to save lives more effectively, by enabling each fire-fighter to become better aware of their group-members' surroundings. Such distributed perceptual systems would require more robust detection mechanisms, and better inference algorithms to incorporate mutually dependent streams of information. On the detection front, I am interested in exploring fusion of multiple sensor-modalities [5] to facilitate robust detection of a richer set of activity-descriptors. Furthermore, I want to extend my existing work on human activity representation [1] [3] to model parallel streams of activity-descriptors for characterizing group-behaviors.

Our long-term research goal is to formalize a basic set of computational models essential for the perceptual understanding of human behaviors. Achieving this goal will advance the study of how we perceive the world around us, and will have a far-reaching impact on such fields as robotics, cognitive science, and ubiquitous computing.

## References

- [1] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, G. Coleman, "Detection and Explanation of Anomalous Activities: Representing Activities as Bags of Event  $n$ -Grams", *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.
- [2] R. Hamid, S. Maddi, A. Johnson, A. Bobick, I. Essa, C. Isbell, "Discovery and Characterization of Activities from Event-Streams", *Conference on Uncertainty in Artificial Intelligence*, 2005.
- [3] R. Hamid, S. Maddi, A. Bobick, I. Essa, "Discovery and Structure From Statistics - Unsupervised Activity Analysis using Suffix Trees", *IEEE International Conference of Computer Vision*, 2007.
- [4] A. Dey, R. Hamid, C. Beckmann, D. Hsu, I. Li, "a CAPPELLA: Programming by Demonstration of Context-Aware Applications", *ACM SIGCHI, Conference on Human Factors in Computing Systems*, 2004.
- [5] R. Hamid, A. Bobick, A. Yezzi, "Audio-Visual Flow - A Variational Approach to Multi-Modal Flow Estimation", *IEEE International Conference on Image Processing*, 2004.