

## S&T Campaign: Information Sciences

### System Intelligence and Intelligent Systems

Heesung Kwon, [heesung.kwon.civ@mail.mil](mailto:heesung.kwon.civ@mail.mil), (301) 394-2501  
 Clare Voss, [clare.r.voss.civ@mail.mil](mailto:clare.r.voss.civ@mail.mil), (301) 394-5615

## Research Objectives

- Identify and develop best methods for enhancing situational awareness through joint Natural Language (NL) Text and Video Analytics
- Leverage advances in extracting semantic meaning from NL & recognizing objects and activities in images/video, as well as methods that apply to these areas jointly for:
  - NL summarization of video
  - Visual question-answering
  - Ontology-supported activity recognition
- Develop multimodal representation of event semantics



Joint text & image analytics can lead to richer content exploitation. For example, French text in sign could id any French-speaking country, while distinctive architecture in image suggests Paris as location. Development of formal methods linking the two is still in its infancy.

## Challenges

- Inadequate data resources for developing algorithms for technology suited to tactical environments
  - Limited, intermittent bandwidth
  - Video collection from robot perspective
  - Noisy, incomplete data
- Difficult to distinguish among contexts & similar activities

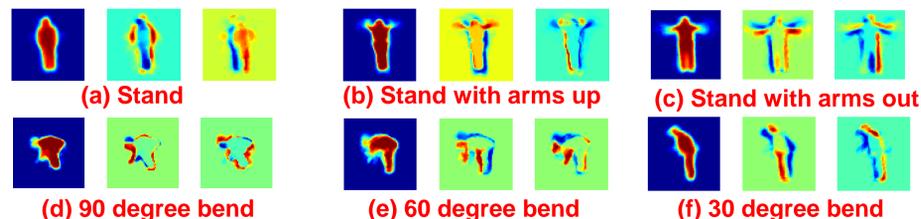


Contextual information is needed to distinguish everyday events (above left) from anomalous, noteworthy events (above right).

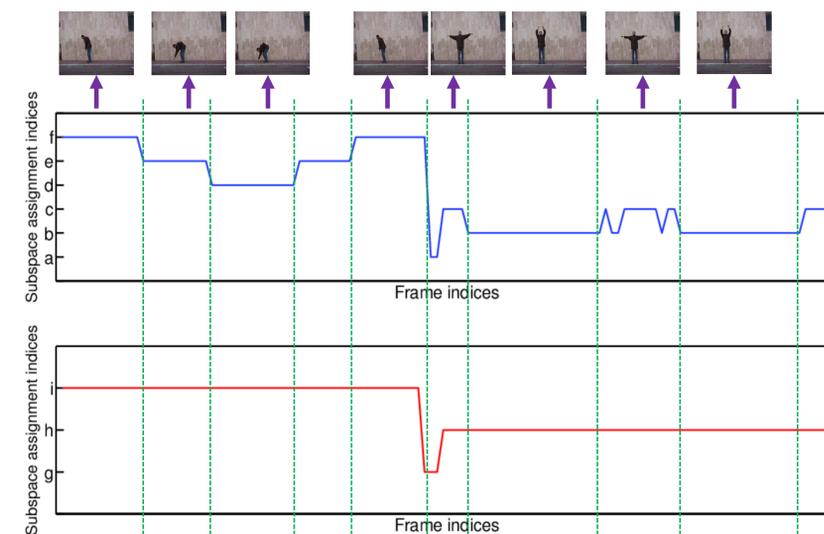
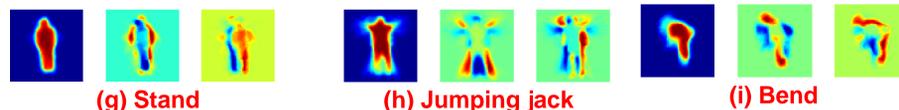
## ARL Facilities and Capabilities Available to Support Collaborative Research

- Facilities and custom-built robots for collecting video data from “first-person” robot perspective
- Resources for crowdsourcing NL text annotations of video and image data to create unique text/video datasets
- Expertise in event ontologies and their links to large-scale NL semantic resources and annotations to support cueing “context” in ambiguous video and image data
  - Ontologies can support breaking down compound events into smaller, more easily recognizable activities (T. Wu, P. Gurrarn, R. Rao, and W. Bajwa, “Hierarchical Union-of-Subspaces Model for Human Activity Summarization,” *Proc. of the IEEE ICCVW on Video Summarization for Large Scale Analytics*, 2015.)
- Specialized focus on recognition and interpretation of events in text and video, as opposed to objects

### Attributes at a finer resolution



### Attributes at a coarser resolution



Ontologies can be used to label & break apart compound actions into recognizable, measurable attribute components at different granularities. Coarse-resolution motions detected in (g), (h), (i) are shown in bottom red graph. These align to finer-resolution motions detected in (a)-(f), shown in upper blue graph.

## Complementary Expertise/ Facilities/ Capabilities Sought in Collaboration

- Seeking expertise in...
  - Natural Language Processing and Computer Vision
  - Multimodal event/activity recognition
  - Machine Learning
- Opportunities to collaborate on unique types of text/video data annotation and analysis