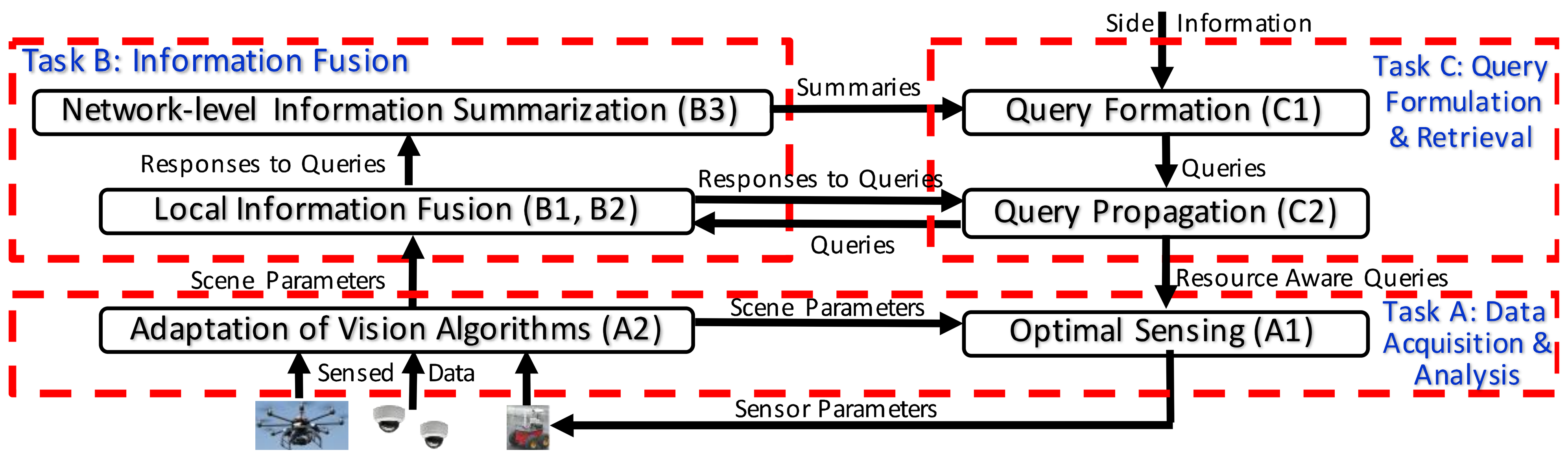
CPS: Synergy: Collaborative Research: Extracting time-critical situational awareness from resource constrained networks

PIs: Amit Roy-Chowdhury, Srikanth V. Krishnamurthy, Eamonn Keogh

Research, Education and Outreach Objectives

- **Research Objective:** Facilitate timely retrieval of situational awareness information from rich content (including video) generated by field deployed nodes in resource-constrained, uncertain environments

- **Major research tasks:**



A. Resource-Constrained Data Acquisition and Analysis
1. Optimally and dynamically reconfigure the activation of field deployed agents to capture relevant information
2. Develop strategies to adapt video analysis algorithms based on environmental conditions and available resources

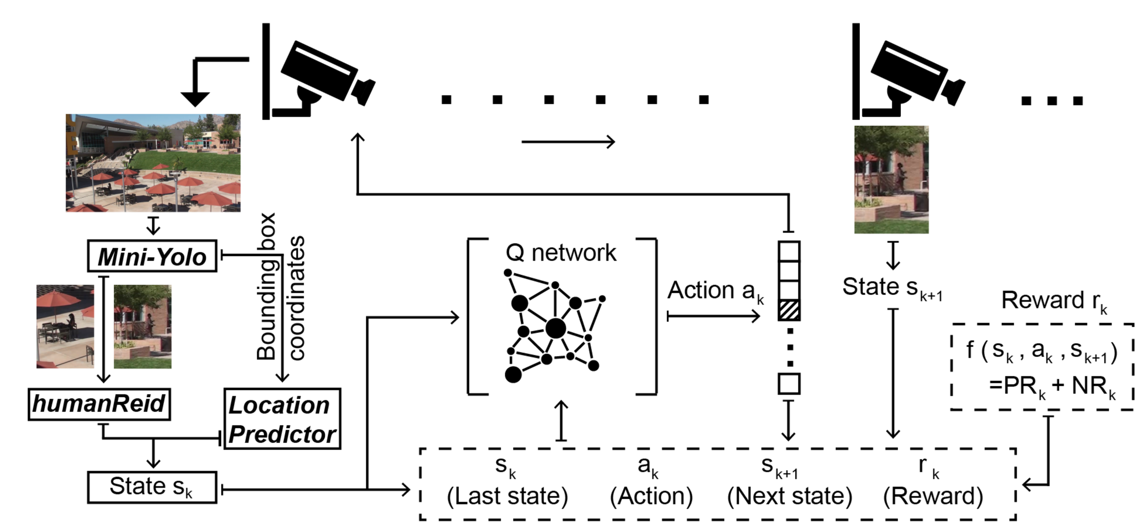
B. Information Fusion Under Resource Constraints
1. Locally process data, estimate its utility and decide what to transmit
2. Fuse data in a distributed manner while accounting for the directional nature of video sensors and the constraint of limited resources
3. Summarize the incoming information at the central station in the presence of missing data

C. Progressive Approach to Scalable Big Data Processing
1. Express queries using a query budget to restrict the amount of data processing time
2. Incorporate query-time analysis to reduce the redundancy of computation
3. Create an efficient workflow of user defined functions to maximize quality of an answer set within the query budget

- **Experimentation:**
  - Extensive experimentation on UCR/UCI camera network testbeds
- **Education and Outreach**
  - Develop specialized graduate and undergraduate courses at UCR and UCI
  - Make tutorials and workshops on content-aware networking and resource-constrained video analysis publicly available

You're It: Controlling a Steerable Camera for Surveillance using Reinforcement Learning

(Screen shots from our experiments on real world experiments)
We design a PTZ camera control algorithm that quickly identifies dynamically arriving targets and frequently acquires high resolution images of existing targets.

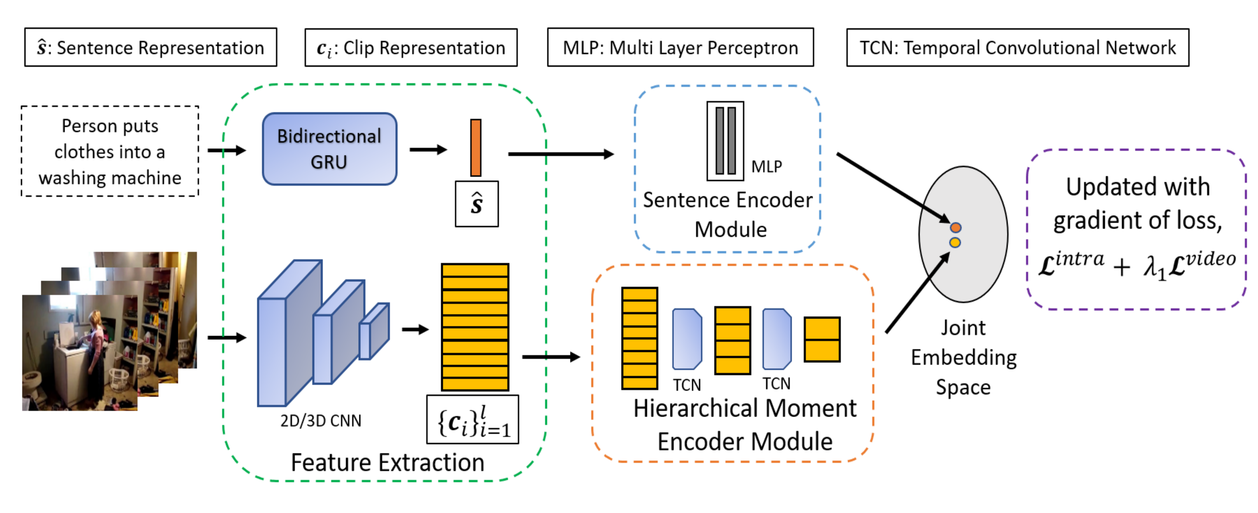


A high-level depiction of our framework

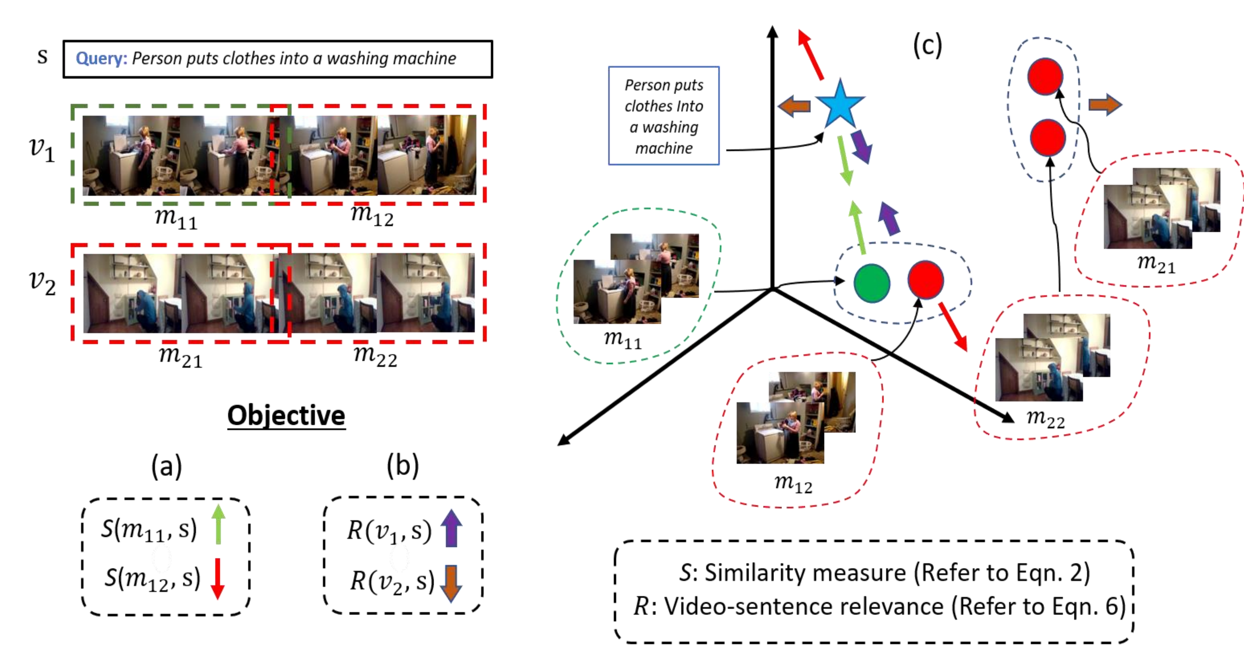
Goal: Balance the tradeoff between (a) zooming out to identify any target quickly when it enters the scene of interest and (b) zooming in on existing targets are frequently as possible to acquire fine grained or high-resolution images to enable tracking of their activities.

Approach:
- The PTZ camera control framework is formulated as Markov decision process (MDP), which selects the PTZ configuration that provide the highest utility in terms of a trade-off between the goals of balancing rapid acquisition of new targets and obtaining fine grained information about existing targets.
- We solve the MDP using Double Deep Q Network (DDQN) (a popular reinforcement learning algorithm)
- Typically, RL agents are training heavy (need heavyweight training to deliver high accuracy). While creation of the requisite, huge number of training instances is possible on fast machines, this is time-prohibitive in our scenario because the PTZ configuration alteration requires mechanical movements where each movement can be of the order of seconds. Hence, we instead create a simulator to mimic the camera, target movements and other dynamics to enable training; we later deploy the trained agent during test time.

Text-based Localization of Relevant Moments in a Video Corpus



Framework illustration of text-based video moment localization from video corpus



Conceptual representation of the proposed learning objective

Goal: Given a text query, identify the corresponding moment in a corpus of untrimmed videos. As a result, the system requires to identify the correct video that the text query belongs to and in the correct video, localize the correct moment by distinguishing intra-video moments based on the nuances of different events.

Approach:
- Our objective is to learn a joint embedding space that will align representations of corresponding video moments and sentences. For this, we propose **H**ierarchical **M**oment **A**lignment **N**etwork (**HMAN**), a novel neural network framework, that effectively learns a joint embedding space to align corresponding video moments and sentences.
- We employ feature extraction units to extract clip level features from videos using 2D CNN/ 3D CNN and sentence features from sentences using bi-directional GRU.
- Temporal convolutional layers are used in a hierarchical setup to project candidate moment representation in the joint embedding space in a single stage approach.
- We design the learning objective to explicitly focus on distinguishing intra-video moments and distinguishing inter-video global semantics.

Publications:

[1] S. Paul, N. C. Mithun, and A. Roy-Chowdhury. "Text-based Localization of Moments in a Video Corpus." arXiv preprint arXiv:2008.08716 (2020).

Collaborative Research with UC Irvine.