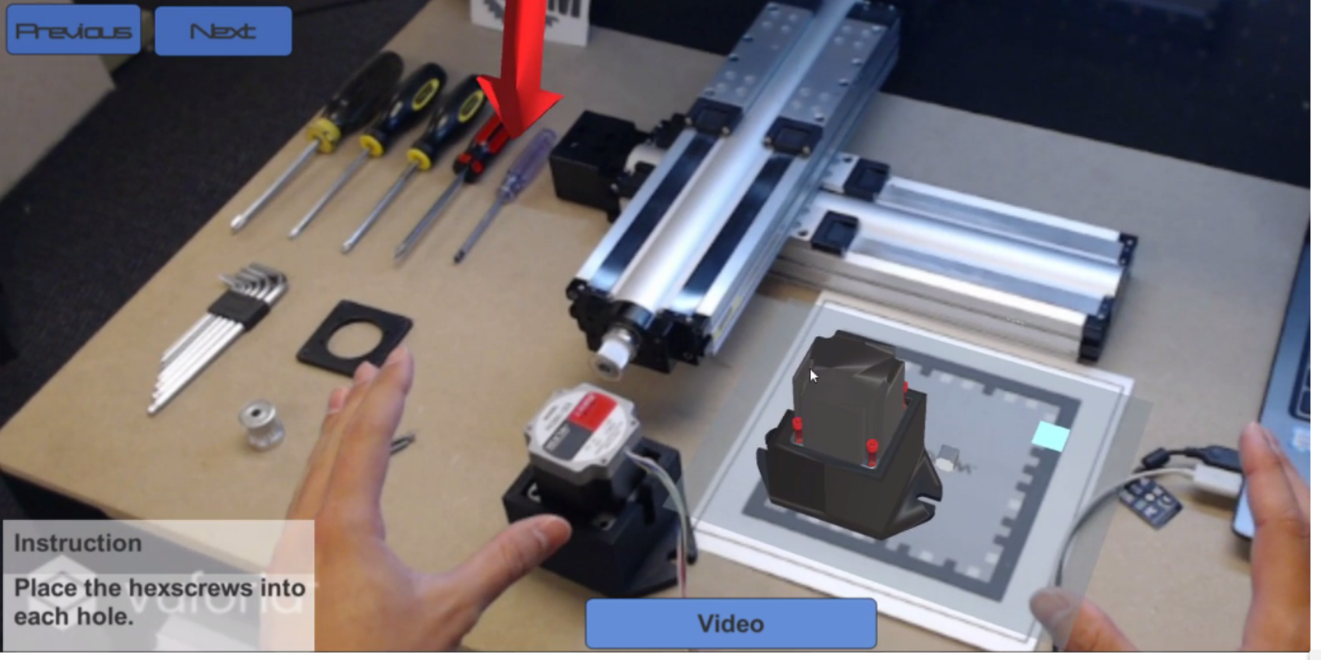


Cyber-Physical Sensing, Modeling, and Control with Augmented Reality for Smart Manufacturing Workforce Training and Operations Management

Students: Md Moniruzzaman[1], Wenjin Tao[2], Ze-Hao Lai[2] , Md. Al-Amin[3] **Investigators : Zhaozheng Yin[1], Ming C. Leu[2], Ruwen Qin[3], Zhihai He[4]**

[1]Department of Computer Science, [2]Department of Mechanical & Aerospace Engineering, [3]Department of Engineering Management & Systems Engineering, Missouri University of Science and Technology; [4]Department of Electrical and Computer Engineering, University of Missouri

Background

While the U.S. manufacturers are investing tremendous efforts and resources to regain the power and growth of manufacturing, especially in the smart manufacturing, they are confronted by a set of critical and challenging issues:

- Lack of workforce with advanced training and skills;
- Need for rapid and individualized training to achieve workforce agility;
- Need for on-the-job personal assistance to improve worker performance and safety.

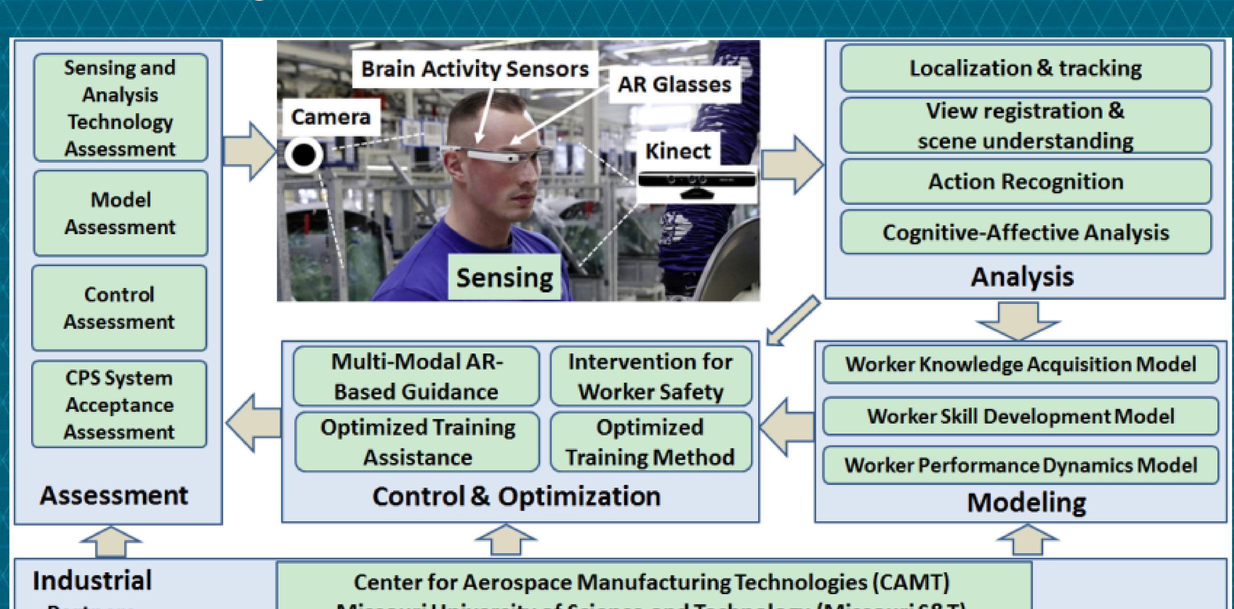
Objectives

This project aims to develop an integrated set of **cyber-physical methods and tools** to sense, understand, characterize, model, and optimize the learning and operations of manufacturing workers, so as to achieve significantly improved efficiency of worker training, effectiveness of behavioral operations management, and safety of front-line workers, for smart manufacturing.

The CPS

A cyber-physical system (CPS) is designed to

- **Sense and analyze** what individual workers are performing, how well they are learning during training or performing tasks in operations, and how safe they are in the workplace, in a quantitative and real-time manner;
- **Model** individual workers learning behavior and operating ability quantitatively, as well as control mechanisms for effectively assisting them, using processed sensor data and retrieved features & patterns;
- **Control** the precise individualized classroom training, adaptive on-the-job training, and multi-modal in-situ task assistance with Augmented Reality(AR), based on the workforce models and real-time sensing.



Research Tasks

- Task I: Cyber-physical sensing and data analytics
- Task II: Data-driven workforce modeling
- Task III: Optimal in-situ assistance to worker training and operations
- Task IV: Performance assessment

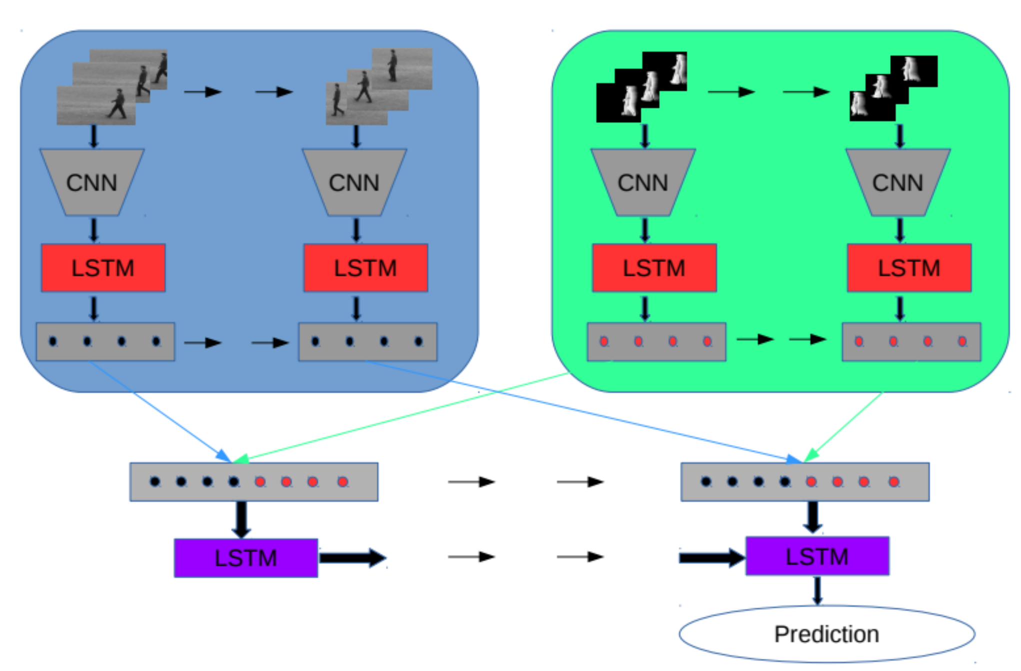
Deep Learning for Worker Action and Behavior Understanding

Automatic understanding of human behavior and interaction with the environment has been an active research area due to its attractive application in a variety of domains. Recognition of human actions through videos is a challenging task, which depends on both spatial and temporal (motion) information.

Two Stream Network

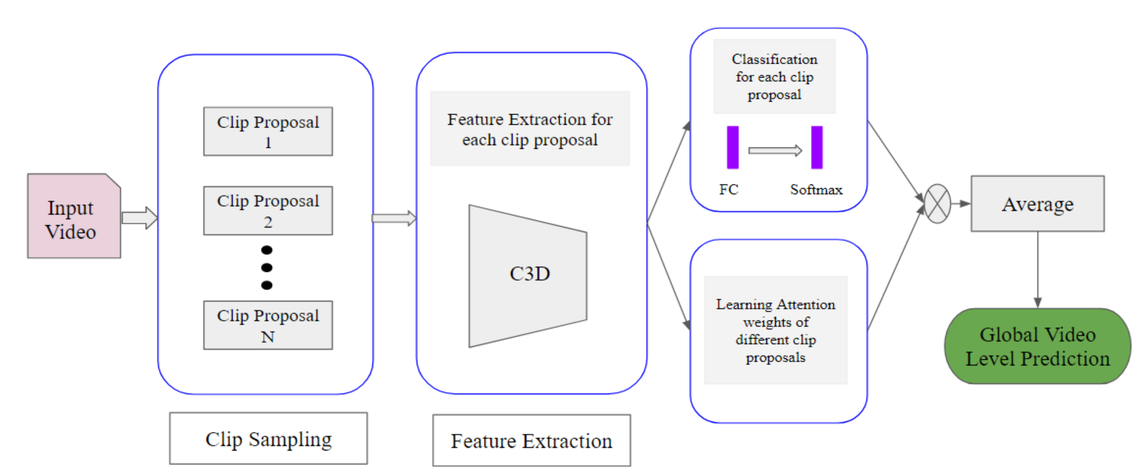
- RGB frames and motion history images were embed into two separate streams.
- Convolutional Neural Networks (CNN) were used for feature extraction and Long Short-Term Memory (LSTM) was used for sequence learning
- Concatenated results obtained from the two stream network were passed through LSTMs and the softmax layer to get the video level prediction.

Two Stream Network



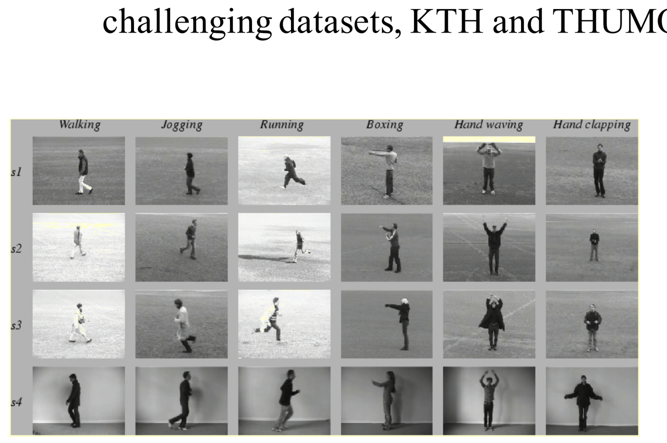
Ranking Model

- A set of attention weights was learned to highlight the discriminative clip proposals and suppress the background clip proposals.



Implementation & Results

- The two stream network and ranking model were evaluated on the public challenging datasets, KTH and THUMOS, respectively.



KTH Dataset

THUMOS Dataset

Method	KTH Dataset (6-actions)
State-of-the-art	95 %
Our two stream model	95 %

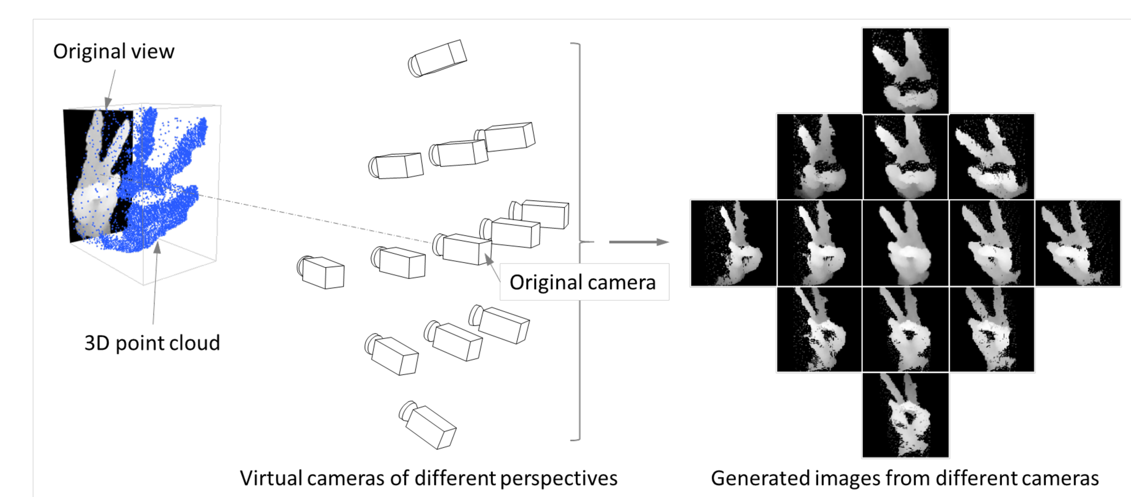
Method	THUMOS Dataset (101-actions)
State-of-the-art	70.75 %
Our ranking model	69.50 % (Top-1 prediction) 83.18 % (Top-2 prediction) 87.39 % (Top-3 prediction)

Hand Gesture Recognition Using CNN with Multiview Augmentation and Inference Strategies

In an engineering assembly task, hand interaction is the major interface between a worker and the work environment. Therefore, hand detection and gesture recognition are crucial for quantification and assessment of worker's performance. However, gesture recognition is still challenging due to the complexities of gestures, high interclass similarity, large intraclass variations and constant occlusions. To address these issues, we proposed the multiview augmentation and inference strategies using Convolutional Neural Network (CNN).

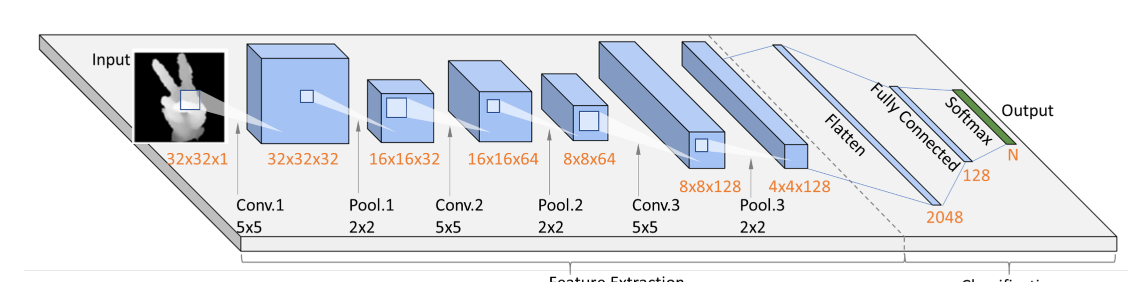
Multiview Augmentation

- Traditional data augmentation methods cannot introduce realistic variations of different perspectives, which are the majority of variations for hand gestures because they are highly perspective-dependent.
- We proposed a multiview augmentation strategy, which generated more data by capturing the 3D point cloud (extracted from original depth image) from multiple perspectives.



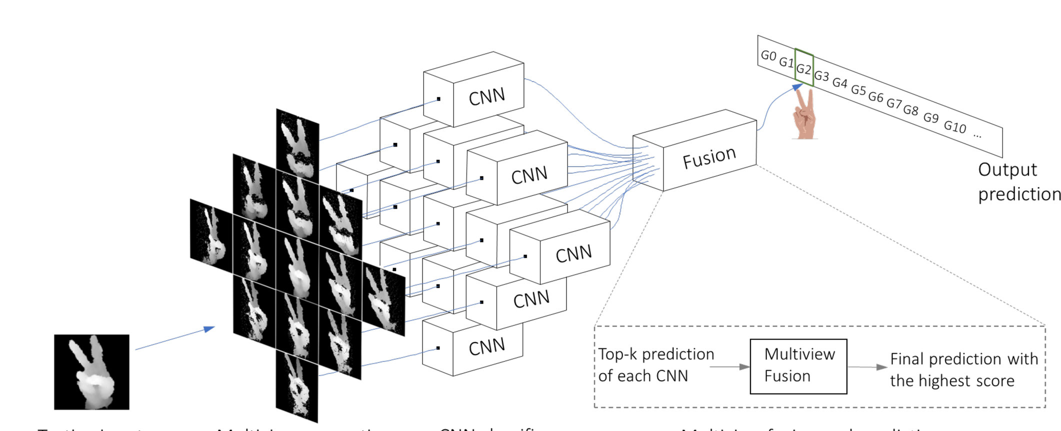
CNN Model for Gesture Classification

- We chose CNN as the basic model to build the classifiern. Our model consisted of a feature extraction module and a classification module.



Multiview Inference

- Inference relying on only one view might not be accurate enough. Therefore, we proposed a multiview inference strategy trying to augment the speculation of each individual view.
  - Every individual view was inferred using the trained CNN.
  - A fusion step was used to find the optimal prediction.



Performance Comparison with State-of-the-arts

- We evaluated our method on two public datasets. Our recognition accuracy outperforms the state-of-the-arts on both of them .

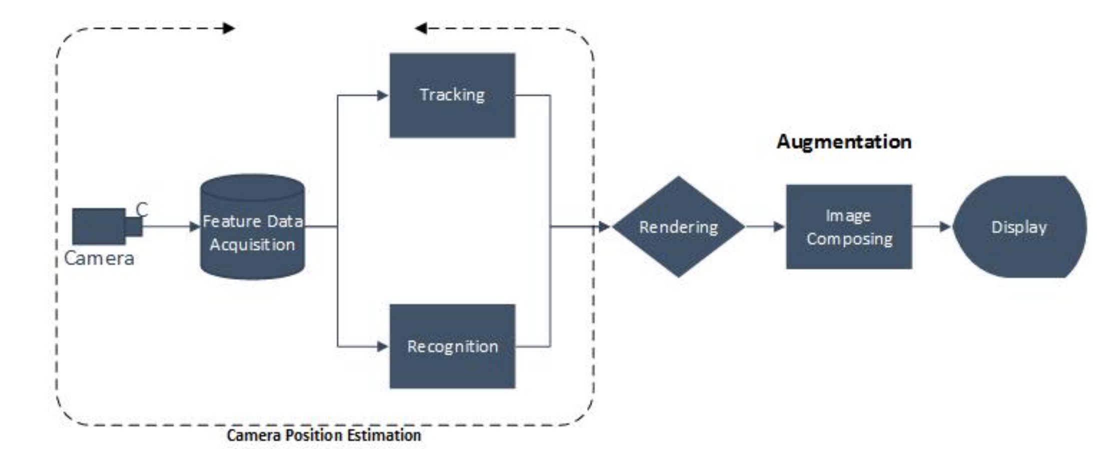
	ASL alphabet dataset (24 gestures, 5 subjects, each has ~500 samples)	NTU 0-9 dataset (10 gestures, 10 subjects, each has 10 samples)
State-of-the-arts	84%	99.6%
Our method	92.7%	100%

Augmented Reality Workforce Guiding System

We developed an Augmented Reality (AR) guiding system for assisting workforce training and operations in smart manufacturing in order to provide a more efficient, effective, and safe working environment.

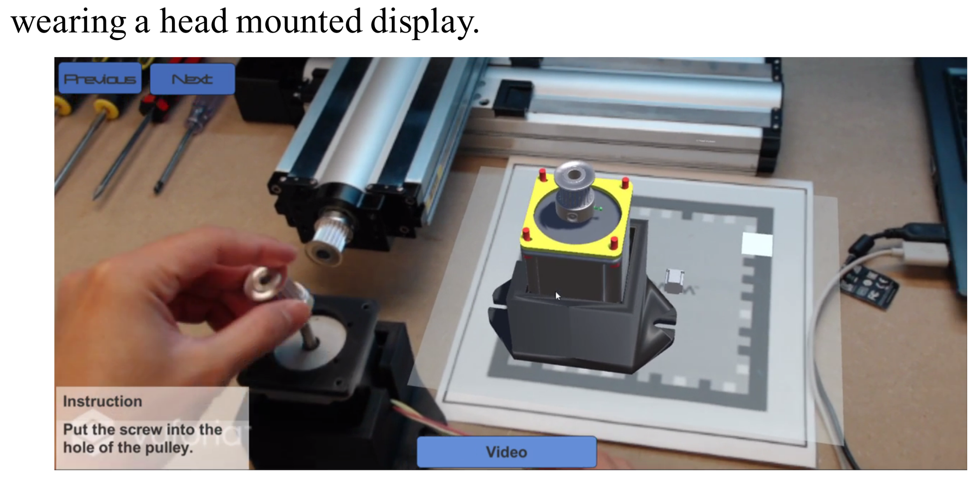
Augmented Reality Principle

- A 3D world coordinate system for augmentation was created based on the marker plane. The relative positions between the camera and markers were obtained using the coordinate transformation matrices.
- Features for augmentation were captured from the fiducial markers.
- After acquiring the features, the overlaying process for augmentation was carried out and augmented objects were displayed upon the 2D camera scene.



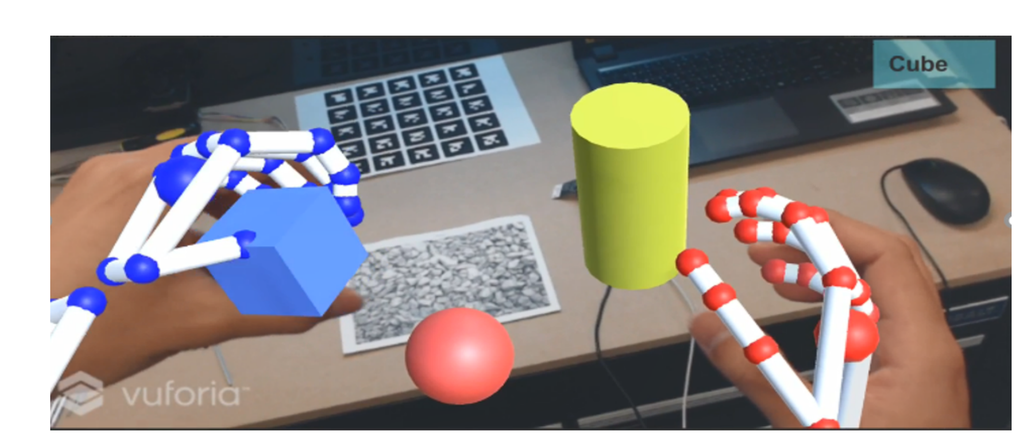
Augmented Reality Assembly Guiding System

- The first-person view composed of AR interface with multi-modal instruction rendering was built, which includes texts, images, annotations, and animations
- The system was tested on the stepper motor assembly process for rail slider mechanism as shown below. The overall task was executed by wearing a head mounted display.



Augmented Reality Based Virtual Object Manipulation

To be able to realize the real-time virtual object manipulation in the workforce training, we leveraged an infrared sensor (Leap motion) and a RGB camera to capture the hand skeleton, then mapped it on to the camera scene blended with augmented objects with different geometries.



Acknowledgements

This research is supported by National Science Foundation grant CMMI-1646162.

NSF MISSOURI S&T