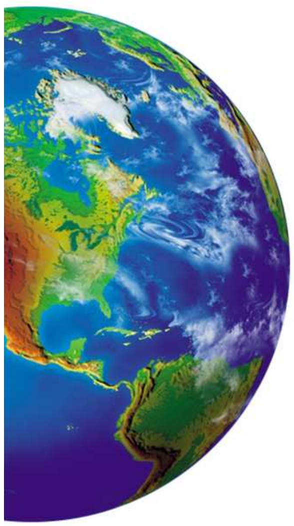


AI Visual Intelligence Technologies for Building a Smarter World

Ming-Ching Chang

University at Albany
State University of New York (SUNY)

The 4th International Conference on Universal Village (UV 2018)

Special Session on

Future Intelligent Manufacturing

Oct. 24, 2018

IEEE

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Ming-Ching Chang

http://www.albany.edu/faculty/mchang2/

- SUNY Albany CS (2018 – present)
- SUNY Albany ECE (2016 – 2018)
- SUNY Albany Adjunct (2012 - 2016)
- GE Global Research (2008 - 2016)
- Brown Univ. PhD (2008)
- National Taiwan Univ. MS/BS (1998)
- ITRI RA (1998)

Bio

Dr. Ming-Ching Chang is an assistant professor at the Department of Computer Science, College of Engineering and Applied Sciences (CEAS), University at Albany, State University of New York (SUNY). He was with the Department of Electrical and Computer Engineering from 2016 to 2018. He was a lead computer scientist in the Visualization and Computer Vision laboratory at the GE Global Research Center in Niskayuna, NY from 2008 to 2016. He was a research assistant in the Laboratory for Engineering Man/Machine Systems (LEMS) at Brown University, where he received his doctoral in Engineering in 2008. He was a research assistant in the Mechanical Industry Research Laboratories, Industrial Technology Research Institute (Taiwan). He received his M.S. in 1998 and B.S. in 1996, both from the National Taiwan University.

Dr. Chang's expertise includes video analytics, computer vision, image processing, and artificial intelligence. His research projects are funded by GE Global Research, DARPA, NIJ, VA, and UAlbany. He is the recipient of the IEEE Advanced Video and Signal-based Surveillance (AVSS) 2011 Best Paper Award - Runner-Up, the IEEE Workshop on the Applications of Computer Vision (WACV) 2012 Best Student Paper Award, the GE Belief - Stay Lean and Go Fast Management Award in 2015, and the IEEE Smart World NVIDIA AI City Challenge 2017 Honorary Mention Award. Dr. Chang serves as the Co-Chair of the AI City Challenge CVPR 2018 Workshop, the Publication and Associate Chair of the IEEE Advanced Video and Signal-based Surveillance (AVSS) 2017, and Co-Chair of the International Workshop on Traffic and Street Surveillance for Safety and Security (IWT4S) 2017 and 2018. He has authored more than 50 peer-reviewed journal and conference publications.

Talk Abstract

- Emerging AI technologies are transforming our world in making everything more convenient, secure, and innovative. Smart mobility devices, automotive, IoT, cloud infrastructures all are pushing the boundary of technology that impacts our daily life.
- How such technology-life transformation can be continuously improved, and how technology can be wisely applied, to ensure the sustainability and quality of life in the long run, is the key issue that the Universal Village is pursuing.
- In this talk, I will introduce how computer vision and AI technologies are making our world smarter in several aspects. Among many changing fronts, I will focus on four topics: (1) smart transportation, which represents the core of smart city with strong technical readiness and impact, (2) video scenario understanding based on human pose estimation, facial analysis, action recognition, and group behavior analysis, (3) industrial video analytics including manufacturing inspection, and (4) digital media forensics that aims to detect fake images and videos. The talk will conclude with technology development trends and future directions.

Talk Overview

- Personal Background & SUNY Albany
- A Smart Future Full of Technologies
  - Brief history of AI
- Computer Vision in Smart Transportation
  - DETRAC Dataset & Benchmark
  - AI City Challenges
  - AVSS T4S Workshop
- Visual Scenario Understanding
  - Human pose, face, group and crowds
- Industrial Video Analytics
- Digital Media Forensics
- Conclusion

University at Albany - SUNY

UNIVERSITY AT ALBANY
State University of New York

https://en.wikipedia.org/wiki/University_at_Albany,_SUNY

From Seeing to Understanding

College of Engineering and Applied Sciences (CEAS)

- Started in 2015
  - was College of Computing and Information in 2005
  - Dept. of CS (25+ faculty), ECE (13+ faculty), Env. & Sus. Engin. (5 faculty). We keep growing.
- New home @ Schuyler building in downtown campus in 2019

Dean Kim Boyer

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Computer Vision and Machine Learning (CVML) Lab

2 faculty, 4 affiliate faculty, 1 postdoc, 8+ Ph.D. students, 8+ alumni, ...

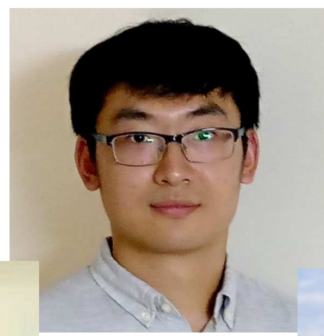
Siwei Lyu
Director

Ming-Ching Chang
Co-Director

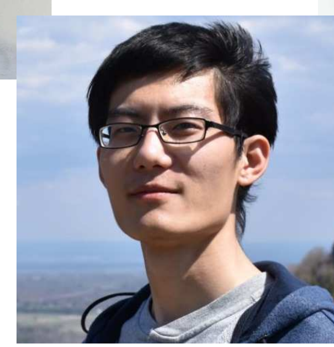
Dawei Do, Ph.D.

Andrew Pulver

Wenbo Li



Yuezun Li



Yi Wei

Lipeng Ke

UNIVERSITY AT ALBANY
State University of New York

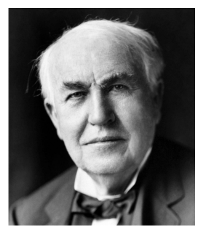
FROM SEEING TO UNDERSTANDING

Who is famous at Albany, the Capital District of New York

NY state gov

GE Research

GE Energy



Thomas Edison

*images in this slide are from the public domain.

...in the research field of Computer Vision and AI

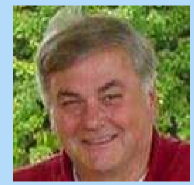
Rensselaer

Rensselaer Polytech Inst.

Kitware

Anthony Hoogs

Joe Mundy



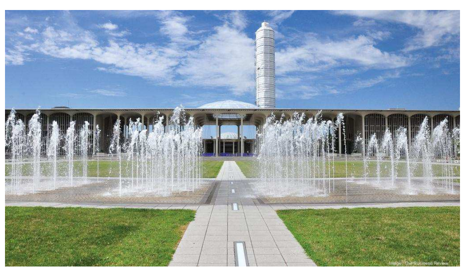
Bill Lorensen

Peter Tu

Richard Hartley

UNIVERSITY AT ALBANY State University of New York

SUNY Office



U. Albany

SUNY Nanotech

Global Foundries

FROM SEEING TO UNDERSTANDING

A Smarter Future Full of Technologies

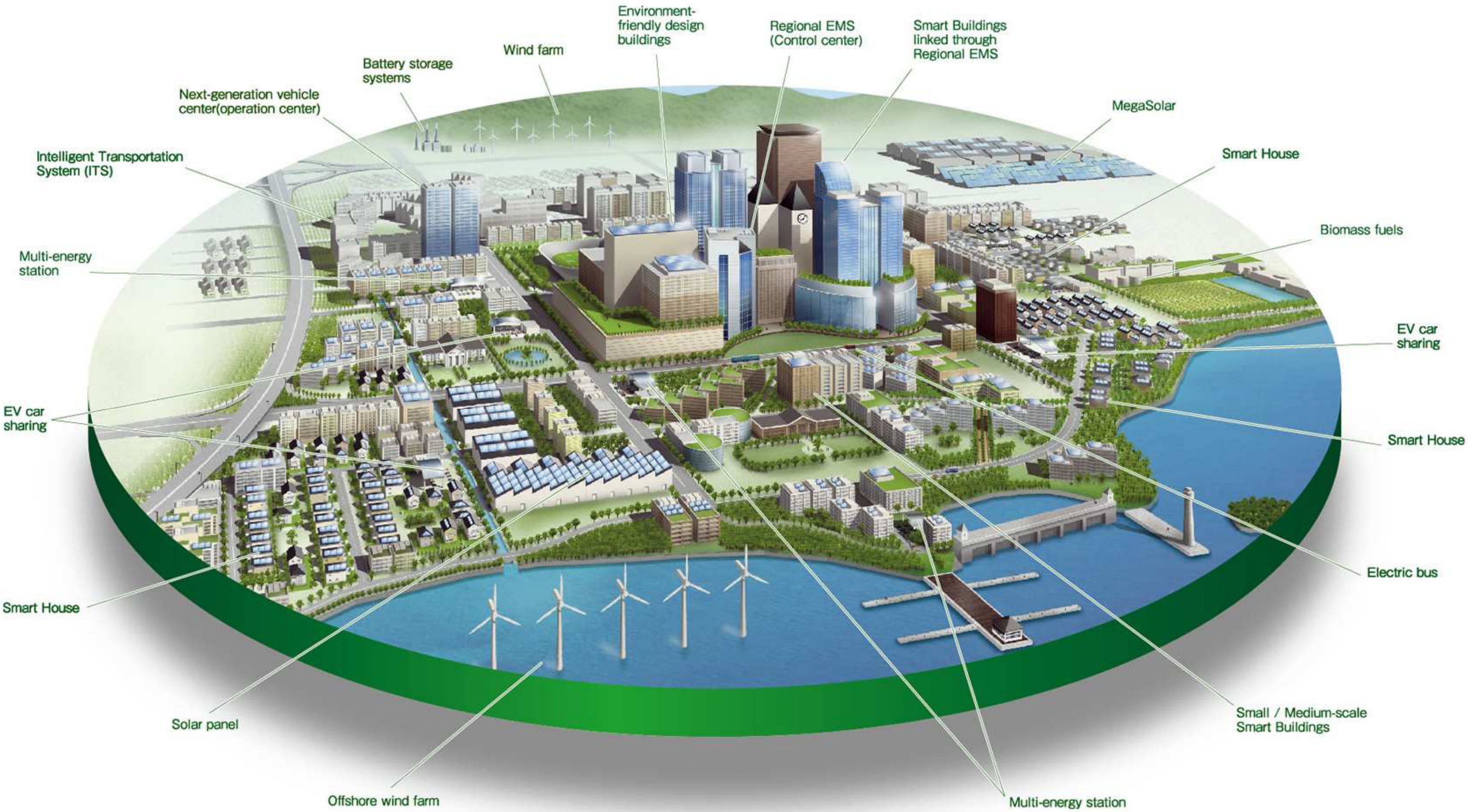


Image from http://www.districtoffuture.eu/

Brief History of AI

THE RISE OF AI

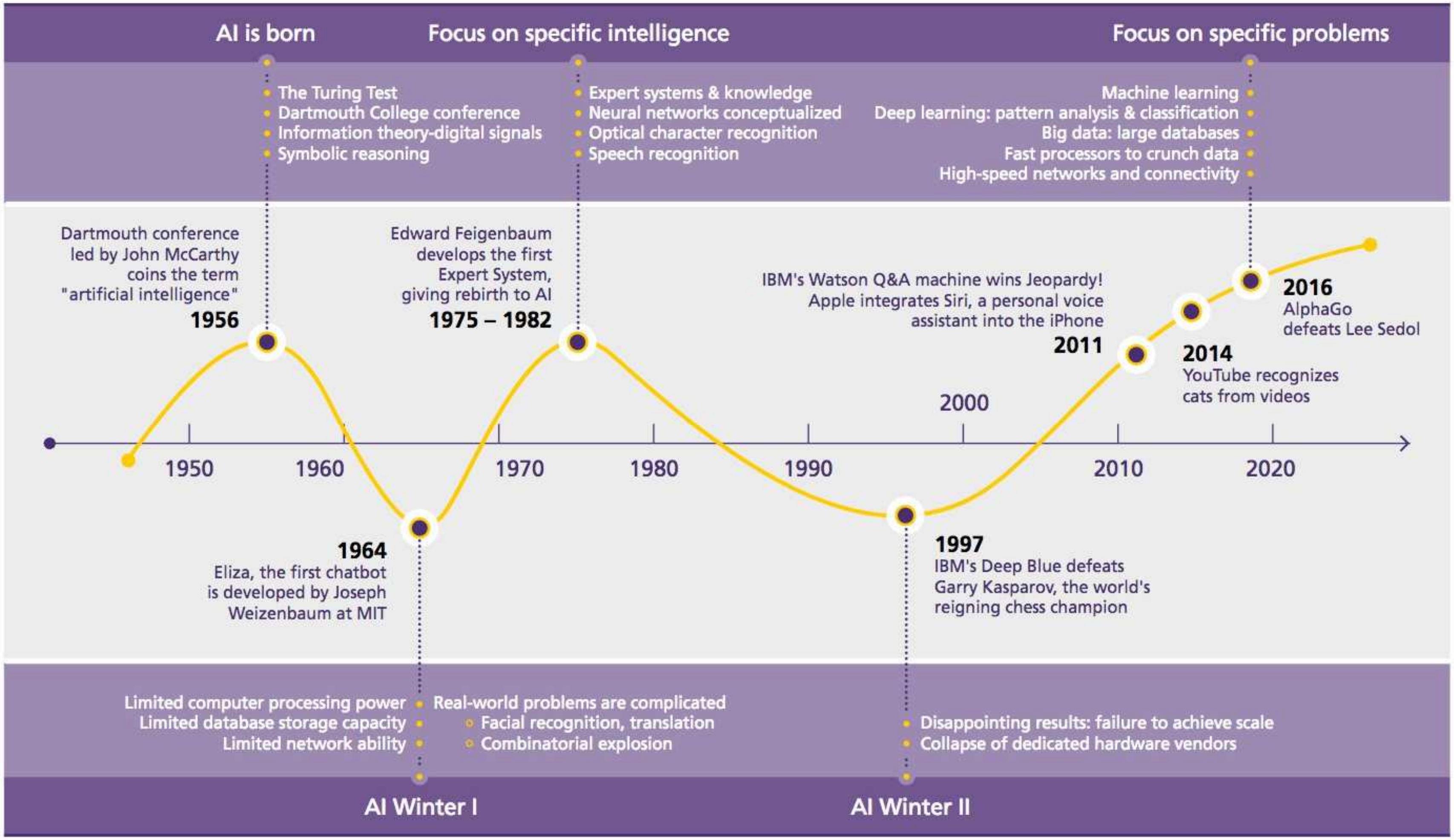


Figure 1: An AI timeline; **Source:** Lavenda, D./Marsden, P.

Alan Turing

Algorithms,

Computation,

Turing machine

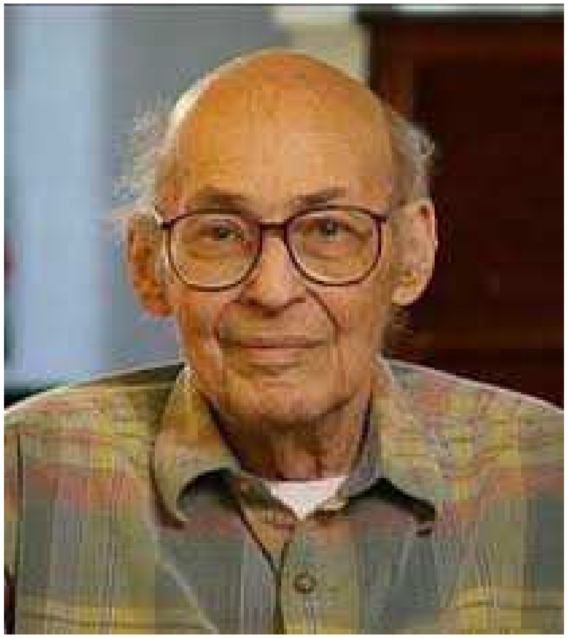
*images in this slide are from the public domain.

Origins of AI

https://avataai.com/origins-of-ai/

- While the concept of artificial systems solving intelligent problems has existed historically, the field of AI research was established at a Dartmouth College conference in 1956.
- Attendees included a de facto list of the Founding Fathers of AI; among them were **John McCarthy, Marvin Minsky, Allen Newell, Arthur Samuel and Herbert Simon**.

Renown pioneers of “classical” AI

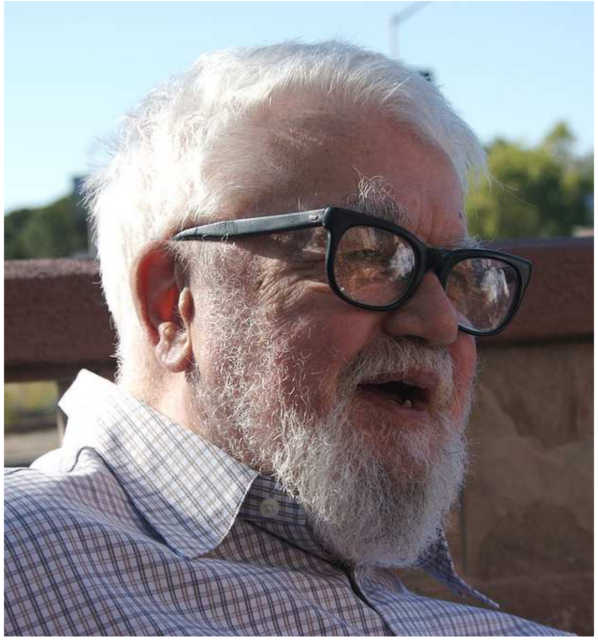


[Marvin Minsky]

MIT

[Herbert Simon]

CMU



[John McCarthy]

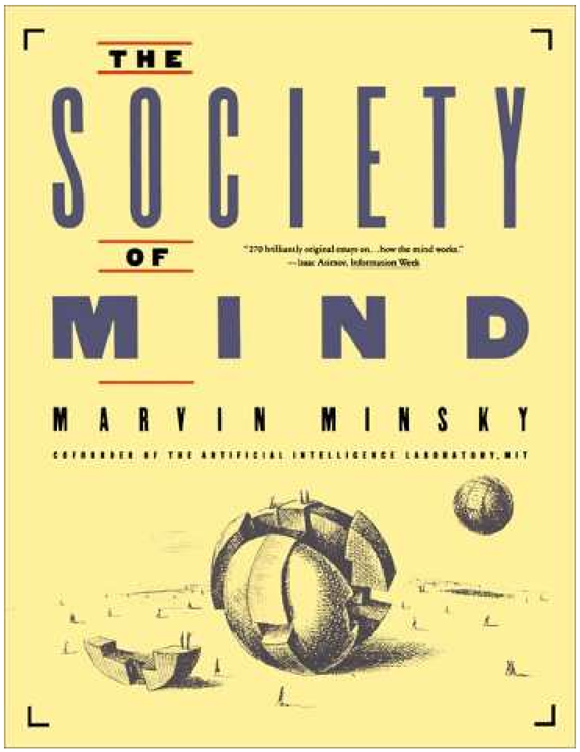
Stanford

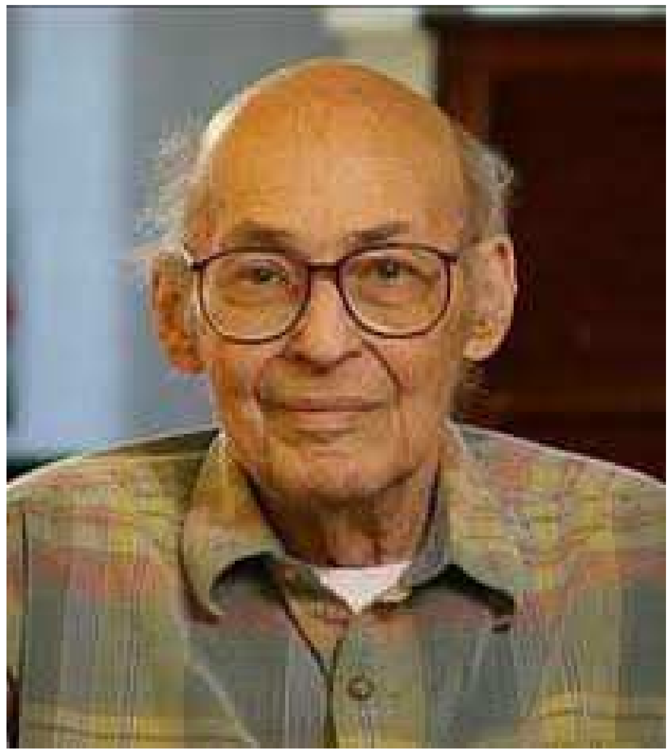
*images in this slide are from the public domain.

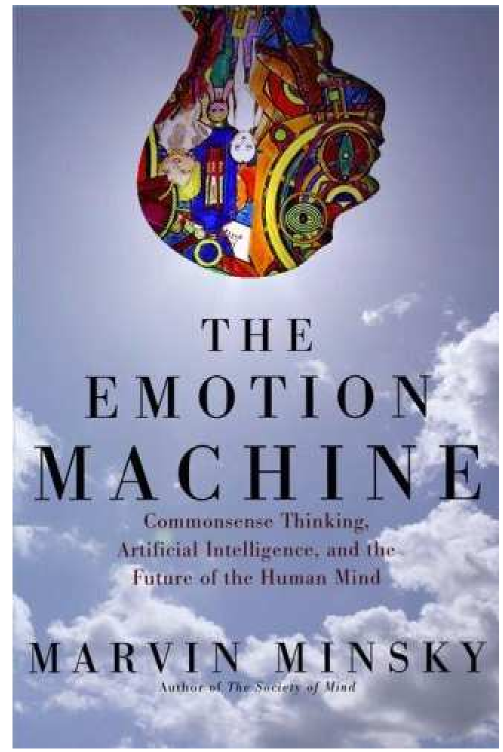
Marvin Minsky

- Wiki: "the farther of AI", co-founder of MIT AI lab

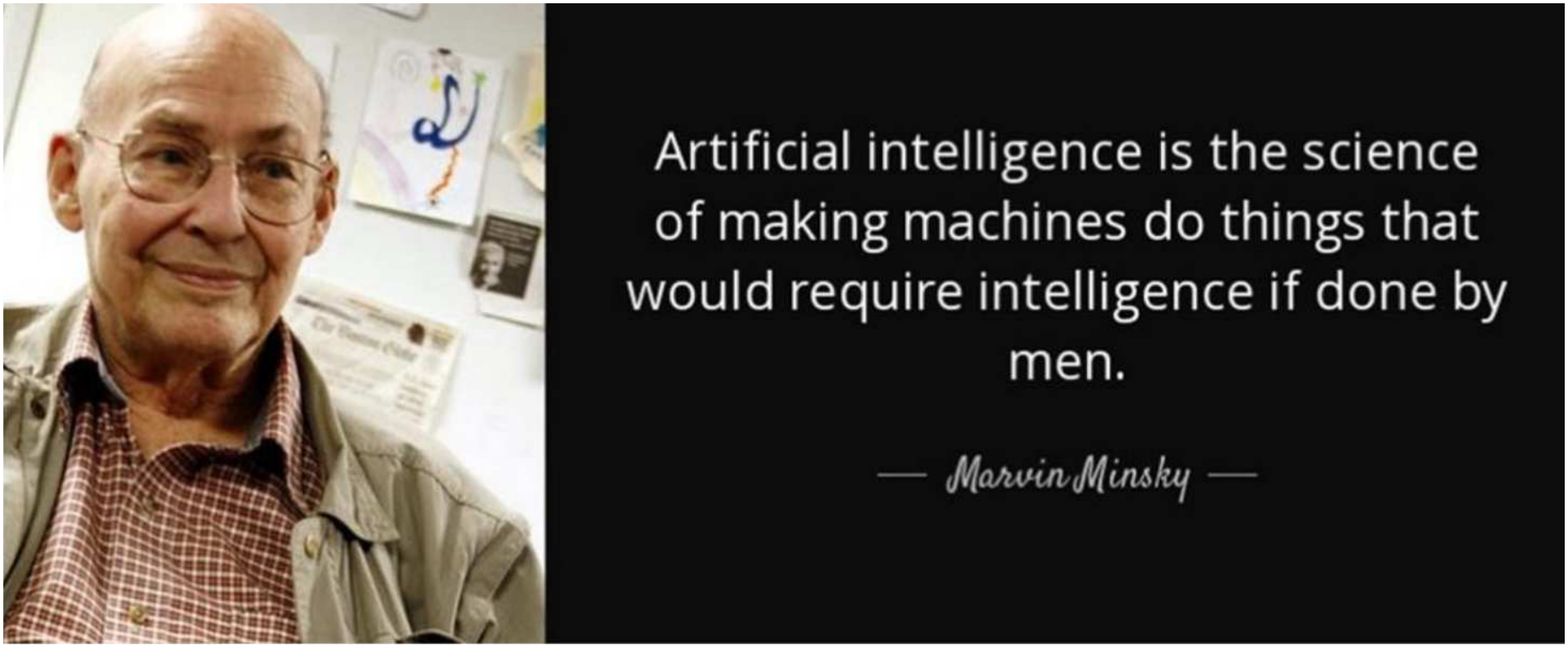
https://web.media.mit.edu/~minsky/







What is AI?


Artificial intelligence is the science of making machines do things that would require intelligence if done by men.
— Marvin Minsky —

The Birth of “Computer Vision”

Slide from Dr. Albert Chen, https://www.viscovery.com/

Why Vision?

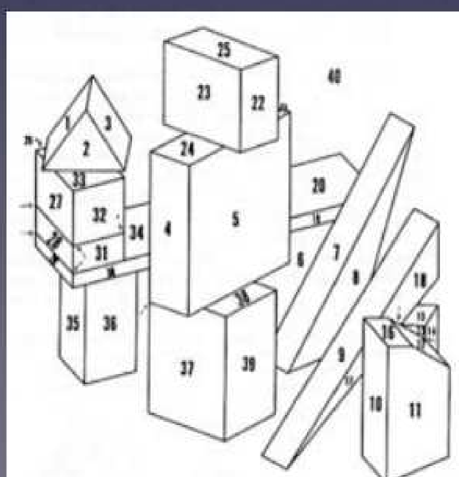
- Prof. Fei-Fei Li [Stanford CS231n] has a good argument
- Evolution's Big Bang
  - The number of animal species grows "exploded" in 543 million years B.C., and why?
  - The first animal developed "eye", and thus the onset of "vision".
  - Once you can see, life becomes much "pro-active" and "intelligent".



Images and contents from Prof. Fei-Fei Li's Stanford CS231n course slides http://cs231n.stanford.edu/.

Brief History of Early Computer Vision

- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts



Guzman '68

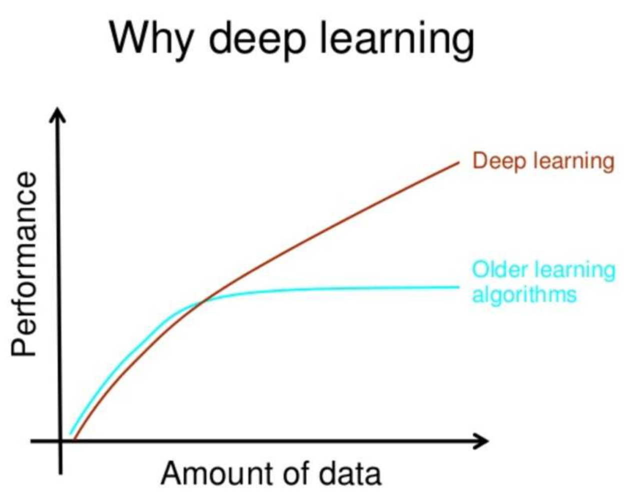
Ohta Kanade '78

Turk and Pentland '91

Slide from Dr. Albert Chen, https://www.viscovery.com/

The Era of Modern AI

- Recent, on-going breakthrough of AI starts as the deep learning (deep neural networks) breakthroughs are taking place
- Modern AI ~= or != Deep Neural Networks



How do data science techniques scale with amount of data?

image from https://blog.algorithmia.com/introduction-to-deep-learning-2016/

Modern Deep Neural Networks

- Comparable (?) or inspired by *biological* (human) neural structures
- Viable based on latest hardware (GPU) developments
- Keep improving in network representations and algorithmic design
- Not seems to reaching its cap yet

Deep Neural Networks (cont'd)

Three *enabling breakthroughs* of DNN:

- Big Data
- GPU
- New algorithmic improvements of DNN
  - basic machine learning techniques
  - new techniques crafted for DNNs: pre-training, dropout, ... etc.

Key Giants in Modern DNN

[Yann LeCun]

NYU
Facebook AI Research

[Geoffrey Hinton]

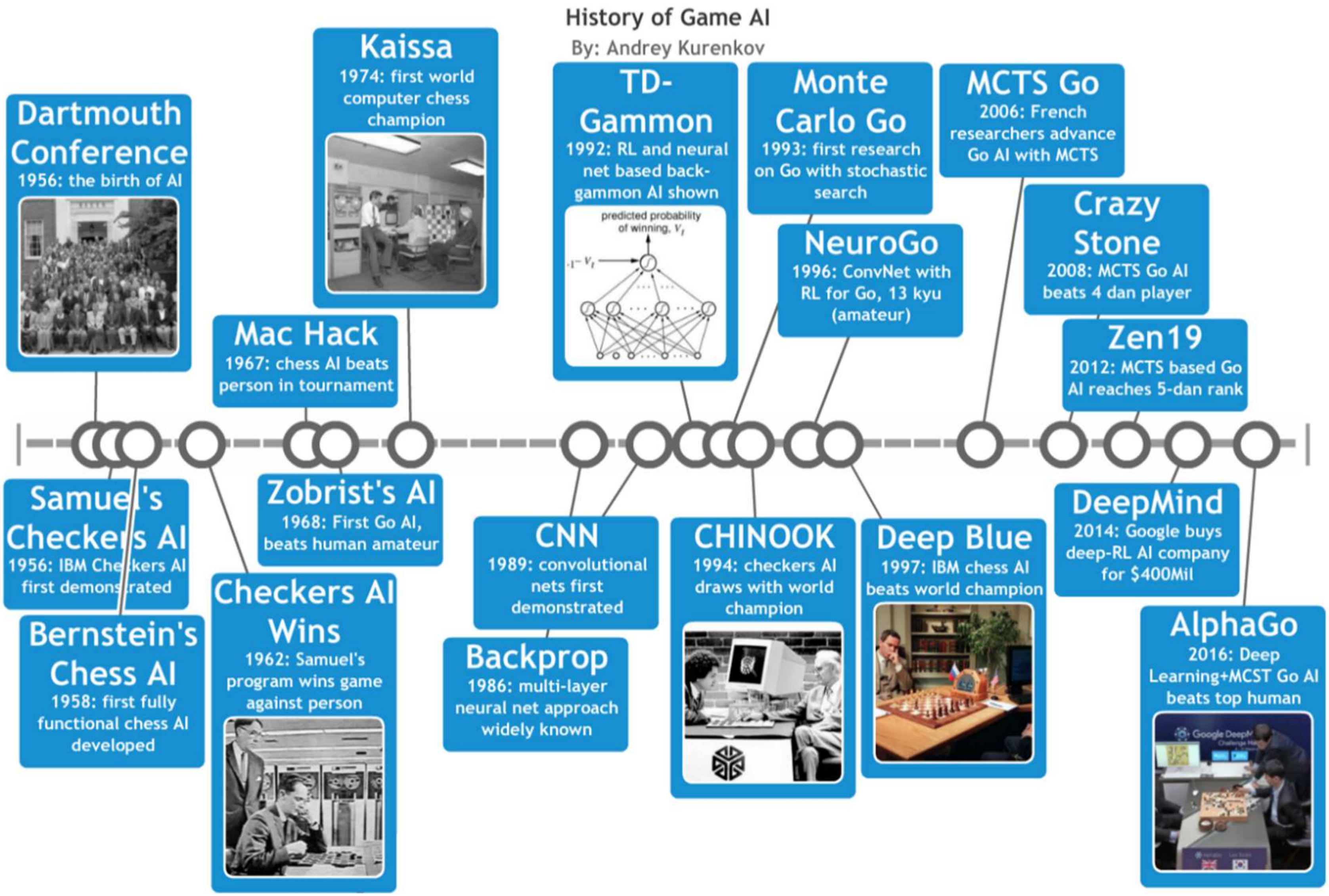
Univ. Toronto
Google Brain

[Yoshua Bengio]

Univ. of Montreal

History of Game AI from Andrey Kurenkov's Blog

http://www.andreykurenkov.com/writing/ai/a-brief-history-of-game-ai/



IBM Deep Blue 1997

From https://en.wikipedia.org/wiki/Deep_Blue_versus_Garry_Kasparov

Deep Blue vs. Kasparov chess

Deep Blue
IBM chess computer

Garry Kasparov
World Chess Champion

First match

- February 10, 1996: takes place in Philadelphia, Pennsylvania
- Result: **Kasparov**–Deep Blue (4–2)
- Record set: First computer program to defeat a world champion in a *classical game* under tournament regulations

Second match (rematch)

- May 11, 1997: held in New York City, New York
- Result: **Deep Blue**–Kasparov (3½–2½)
- Record set: First computer program to defeat a world champion in a *match* under tournament regulations

CPU: 120MHz RS/6000

Technology: game tree search (6-8 moves)

IBM Watson 2011

From https://en.wikipedia.org/wiki/Watson_(computer)

In 2011, the Watson computer system competed on *Jeopardy!* against former winners Brad Rutter and Ken Jennings winning the first place prize of $1 million.

Ken Jennings, Watson, and Brad Rutter in their *Jeopardy!* exhibition match.

Technology: distributed computing, Hardoop cloud

AlphaGo 2015

From https://en.wikipedia.org/wiki/AlphaGo

AlphaGo uses a Monte Carlo tree search algorithm to find its moves based on knowledge previously "learned", specifically by an artificial neural network.

The search is guided by a *value network* and a *policy network*, both implemented using deep neural network.

It was trained further by being set to play large numbers of games against other instances of itself, using reinforcement learning.

Technology: MC Tree Search, Deep Neural Network, Reinforcement Learning

So What's Next?

CMU's **Libratus** that beats pros at Texas Hold'em (2017)

https://www.scientificamerican.com/article/time-to-fold-humans-poker-playing-ai-beats-pros-at-texas-hold-rsquo-em/

AI Real-time Strategy Game (Starcraft) Player

https://github.com/davechurchill/commandcenter

SkyNet??

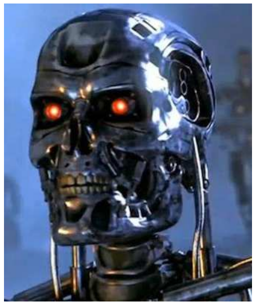


image from https://alumni.berkeley.edu/california-magazine/just-in/2017-08-10/should-we-be-worried-about-rogue-ai-we-ask-experts

AI for a Smarter World

Big Data → Cloud Computing → Edge Computing → IoT → Deep Learning → Fast Network (5G)

Smart agents that can make rational decisions and actions

↓

Artificial Intelligence will be applied everywhere in our live

↓

Image from http://pragativadi.com/chief-secretary-sets-timelines-bhubaneswar-smart-city-projects/

So What Exactly Is AI?

Whenever you see something amazing, always ask yourself “Is this AI”?

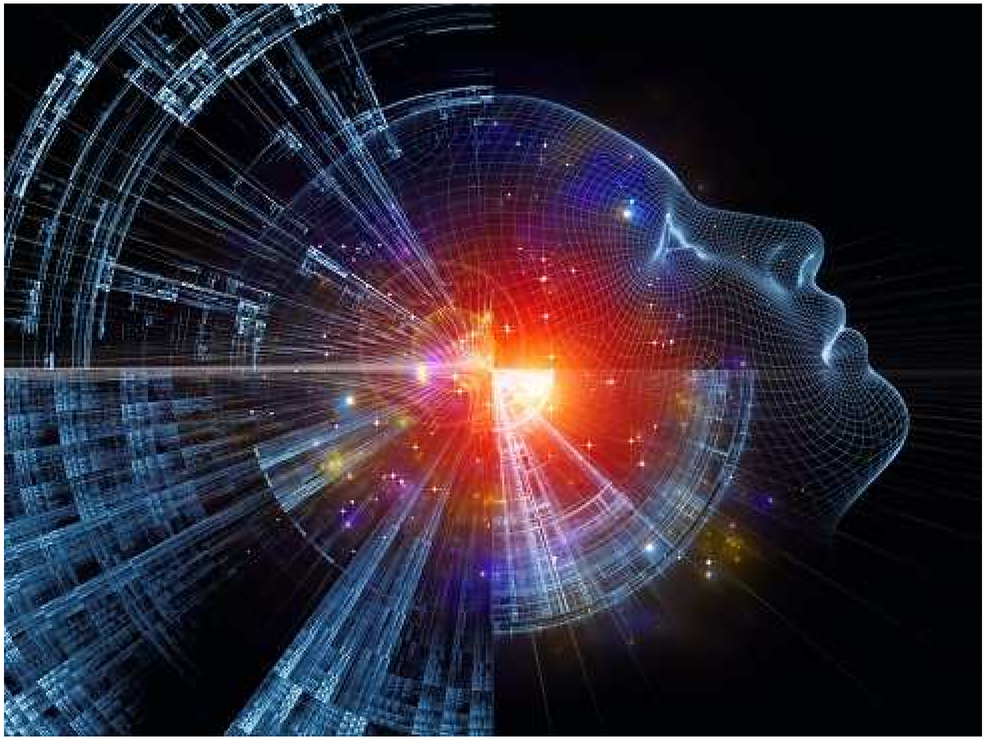
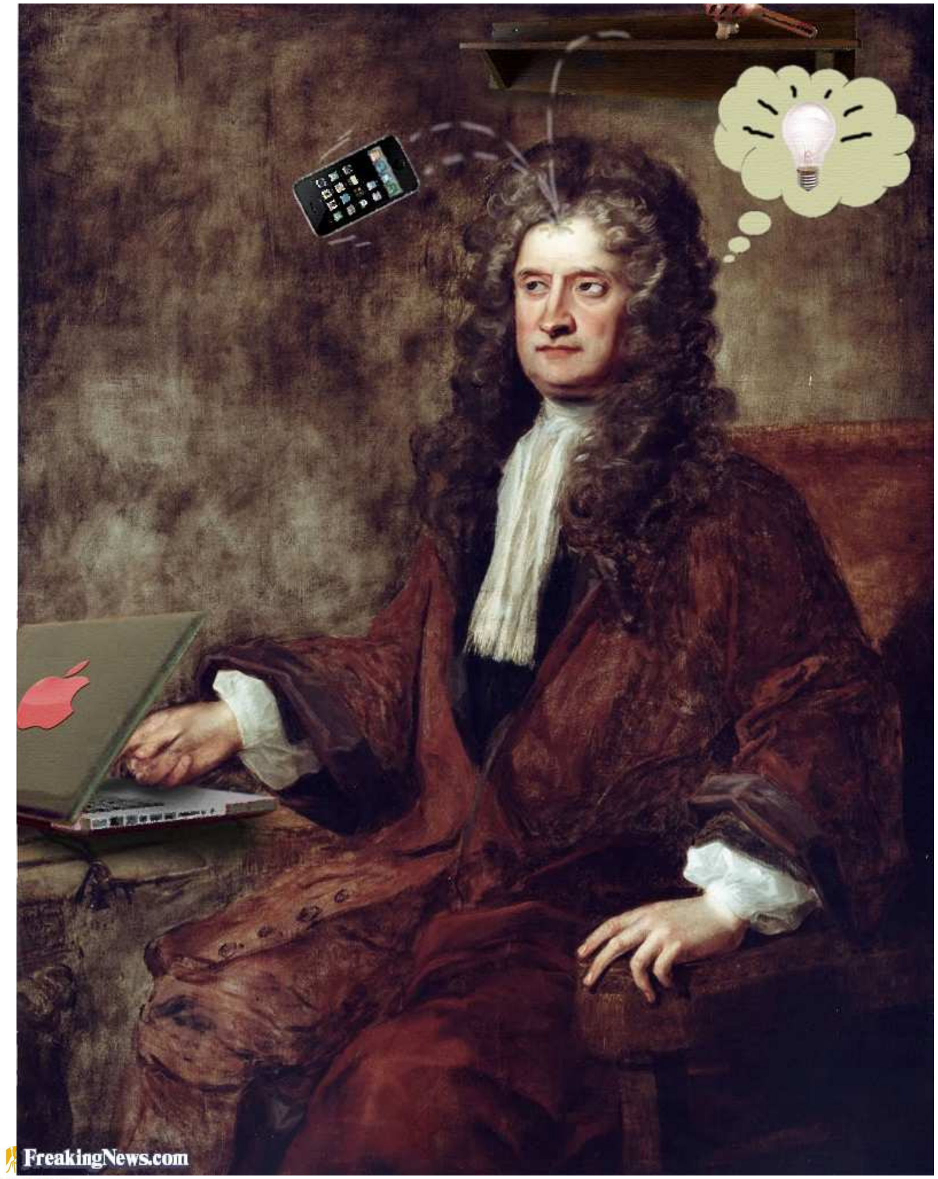


Image from Brain–Computer Interfaces Handbook: Technological and Theoretical Advances, CRC Press

Is this AI?



What if Sir Issac Newton (1642 – 1726) see today's iPhone?

- Note 1: Battery was invented around 1800's (Franklin, Faraday).
- Note 2: Newton is so into "Alchemy" that wouldn't he see the shiny iPhone as some sort of "black magic"?
- Note 3: But don't forget Newton is a smart guy, a "genius", who invented the calculus, law of motion, law of gravitation, ...
- So, is your cellphone some sort of "miracle"?

http://www.freakingnews.com/Isaac-Newton-and-the-Apple-iPhone-Pictures-89960.asp

Smart Transportation

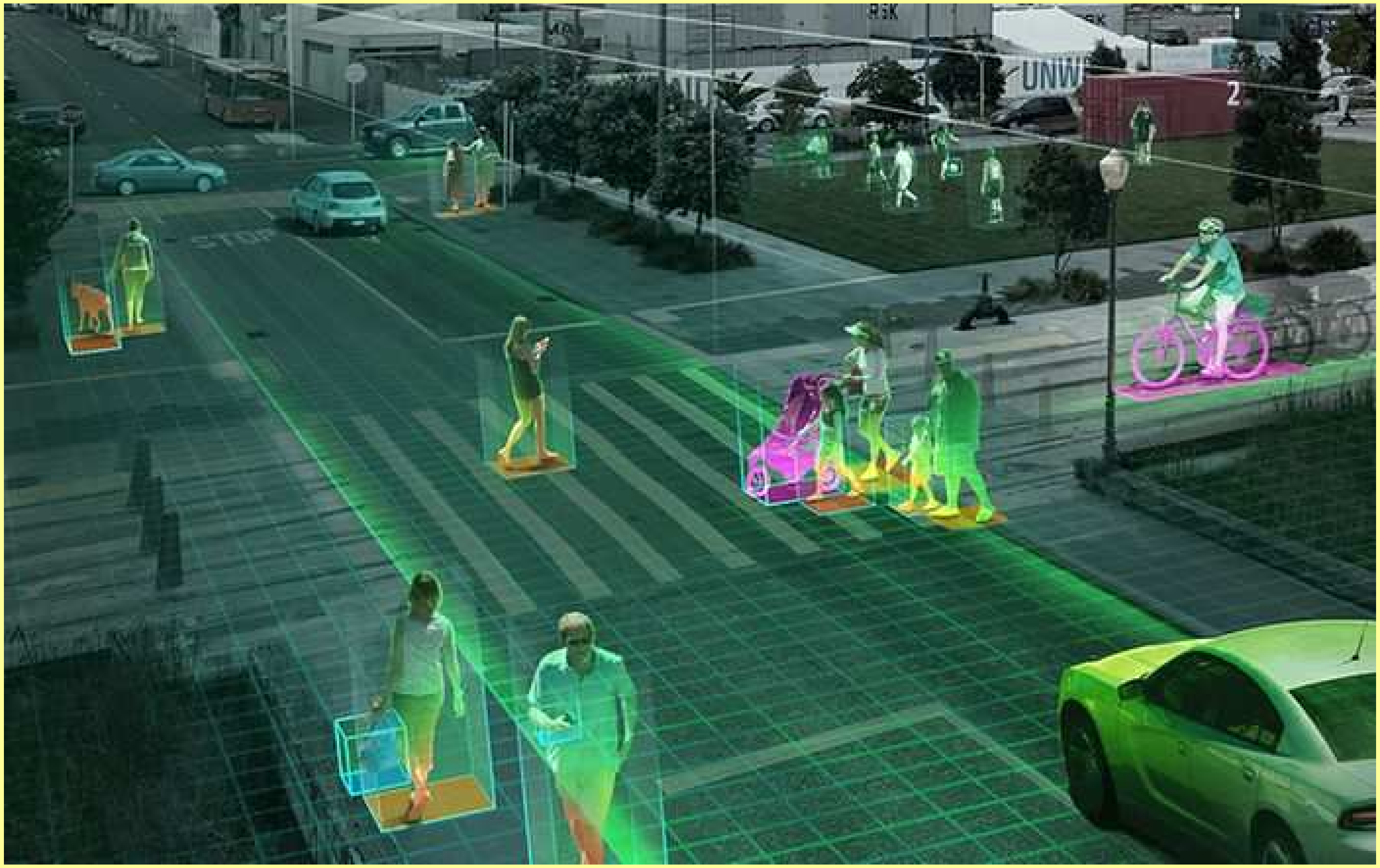


Image from NVIDIA AI City Challenge
https://nvidianews.nvidia.com/news/nvidia-paves-path-to-ai-cities-with-metropolis-edge-to-cloud-platform-for-video-analytics

Smart Transportation

- Smart Transportation is a core of Smart City/World
  - Make public transit systems safer, smarter, and cheaper using ubiquitous street cameras
  - Support strategic decisions for surveillance, safety, traffic control, parking, infrastructure investments, assisting self-driving cars
- Core technology: vehicle/person detection & tracking and analysis

UA-DETRAC Benchmark

University at **A**lbany **DE**tection and **TRAC**king dataset and benchmark

http://detrac-db.rit.albany.edu/

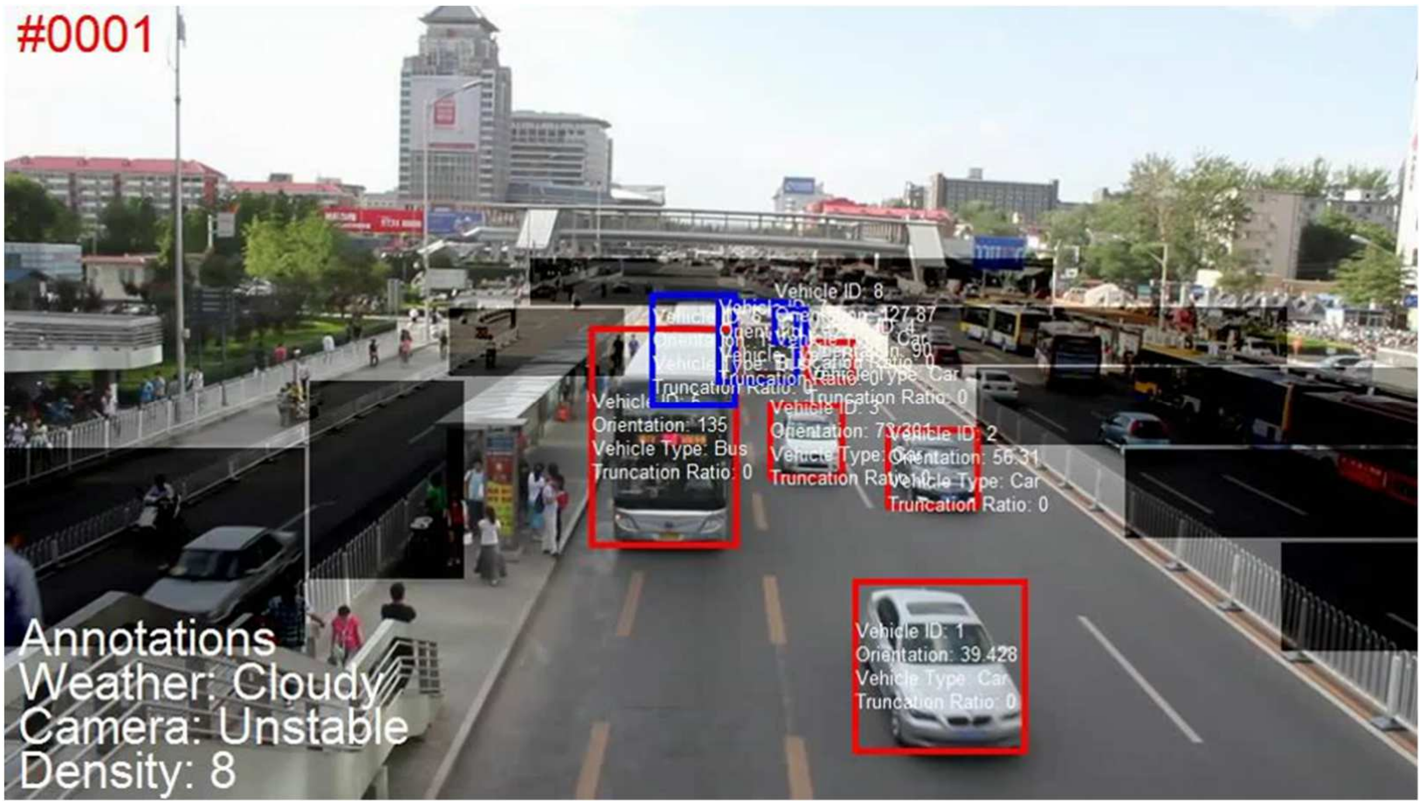
- Large scale vehicle detection and tracking evaluation framework
- Provide *dataset*, *open-src codes*, & *evaluation benchmark*

[L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. "UA-DETRAC: A new benchmark and protocol for multi-object tracking." CoRR, abs/1511.04136, 2015.]

UA-DETRAC Benchmark

University at **A**lbany **DE**tection and **TRAC**king dataset and benchmark

- 100 video sequences of traffic surveillance videos (~1 min. each)
- 140K frames of over 8K vehicles and 1.2M annotated bounding boxes



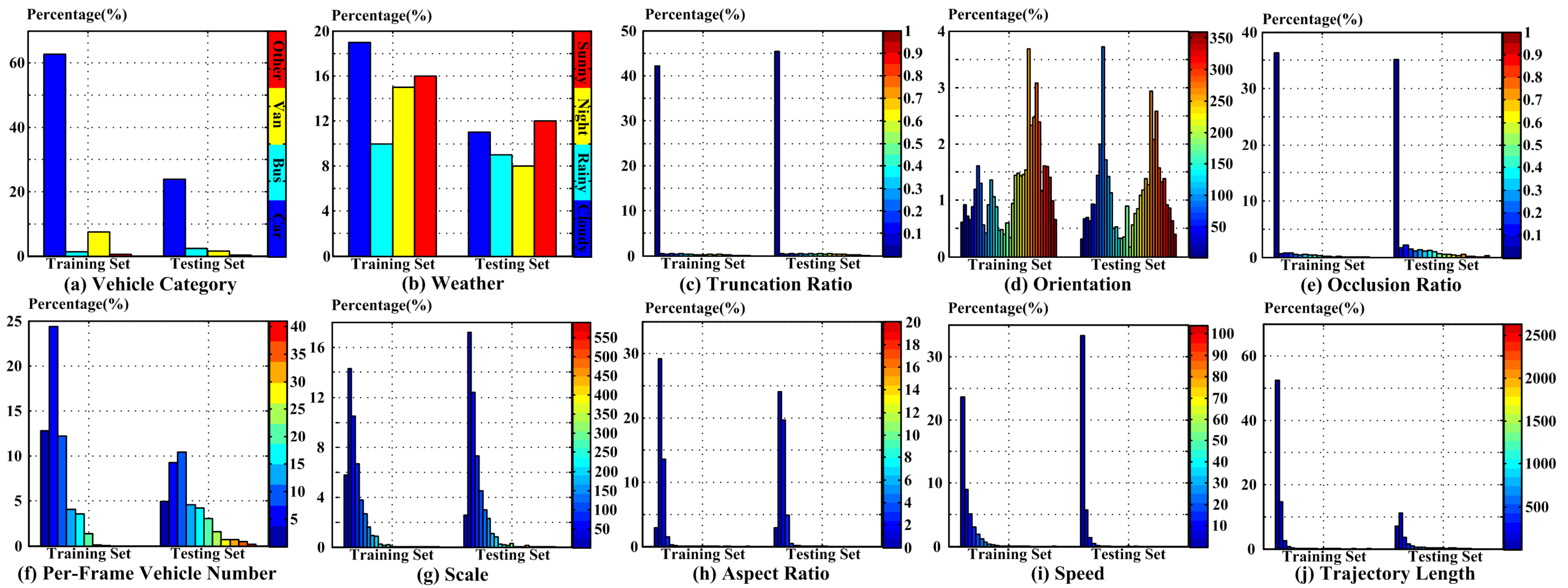
[L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. "UA-DETRAC: A new benchmark and protocol for multi-object tracking." CoRR, abs/1511.04136, 2015.]

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

UA-DETRAC Benchmark

Annotation attributes: weather, vehicle type, orientation, speed, occlusion ratio



[L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. "UA-DETRAC: A new benchmark and protocol for multi-object tracking." CoRR, abs/1511.04136, 2015.]

University at Albany State University of New York

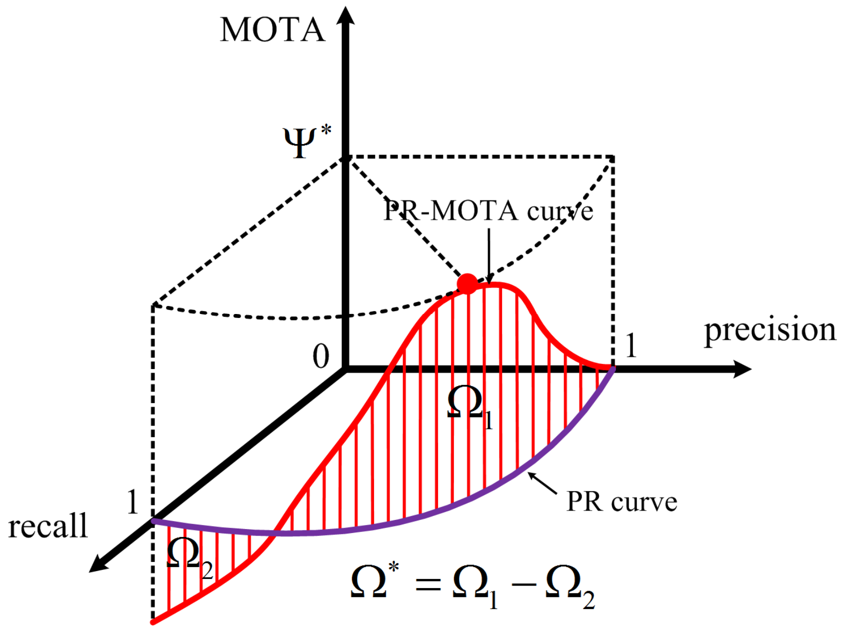
From Seeing to Understanding

DETRAC Evaluation Metric

- Detection: average precision (AP)
- Tracking: DETRAC PR-MOTA

  - Precision Recall
  - Multi-Object Tracking Accuracy

$$MOTA = 1 - \frac{FN + FP + IdSw}{GT}$$



[L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. "UA-DETRAC: A new benchmark and protocol for multi-object tracking." CoRR, abs/1511.04136, 2015.]

AI City Challenge

IEEE Smart World NVIDIA AI City Challenge
August 5, 2017, San Francisco, USA

http://smart-city-sjsu.net/AICityChallenge/index.html

2017

IEEE · NVIDIA · Smart World Congress 2017 · IEEE computer society · IEEE Computational Intelligence Society · TCSC IEEE

2018

CVPR 2018 SALT LAKE CITY • JUNE 18-22

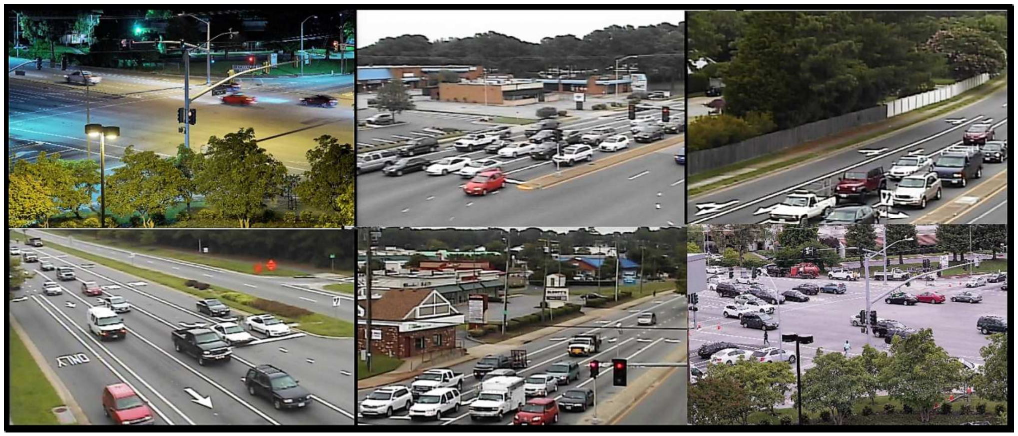
https://www.aicitychallenge.org/

2019

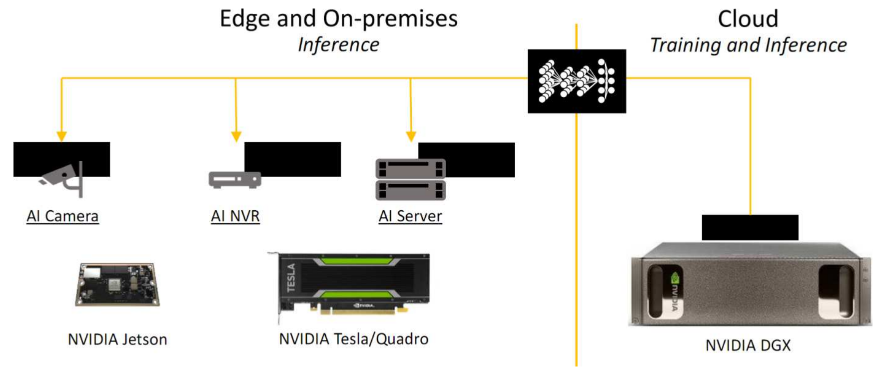
Preparation On Going

2017 NVidia AI City Challenge

Videos captured at Silicon Valley and Virginia



NVidia platform support



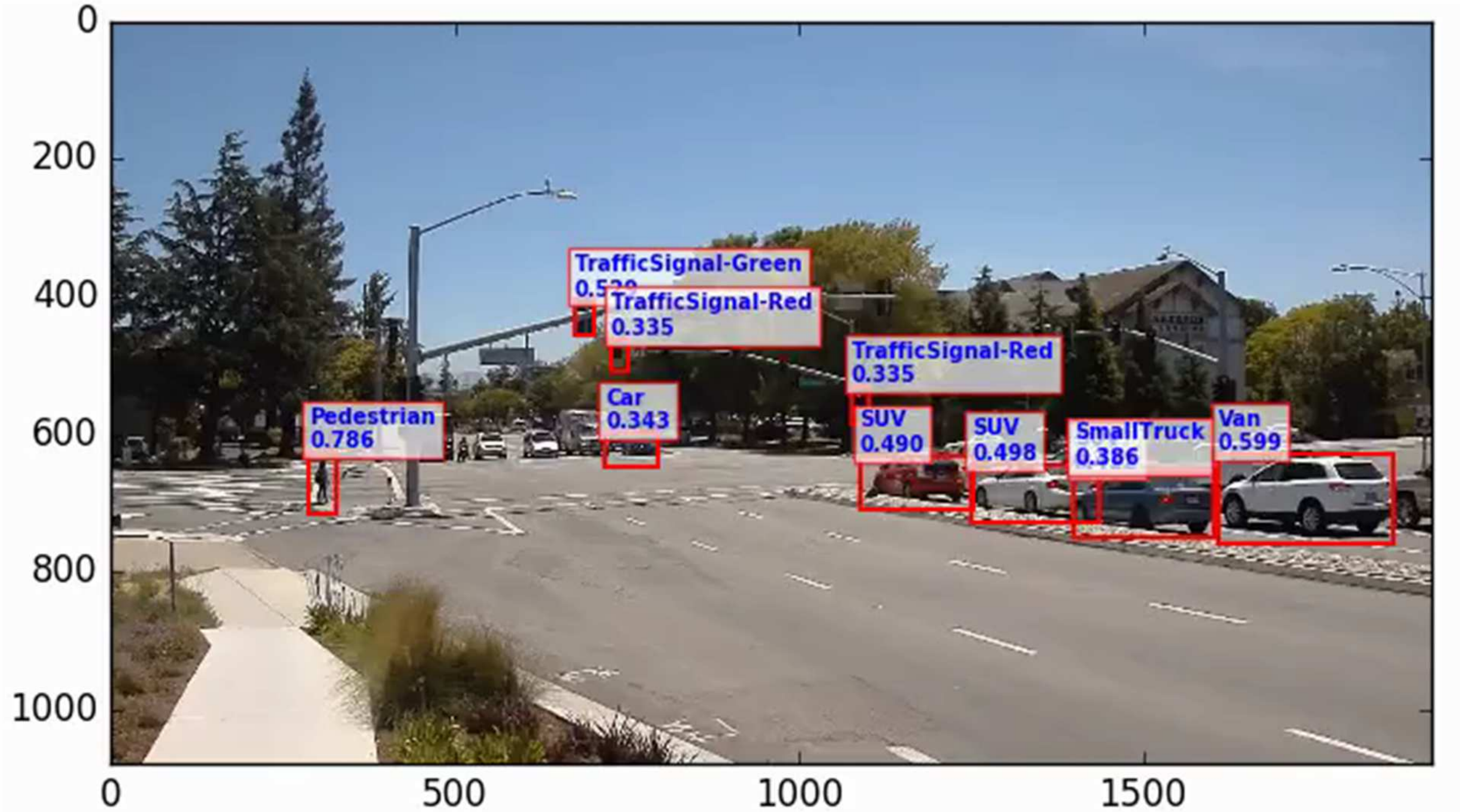
worldwide participation by 29 teams.

- Track 1: vehicle detection
  - winner – UIUC
- Track 2: traffic applications
  - winner – U. Washington
  - honorable mention: UAlbany

[M. Naphade, D. Anastasiu, A. Sharma, V. Jagrlamudi, H. Jeon, K. Liu, M.-C. Chang, S. Lyu, Z. Gao, "The NVIDIA AI City Challenge", IEEE Smart World Congress, 2017]

2017 AI City Detection Test Results

Detection and Classification Results (vehicles, pedestrians, traffic signals, *etc*…)



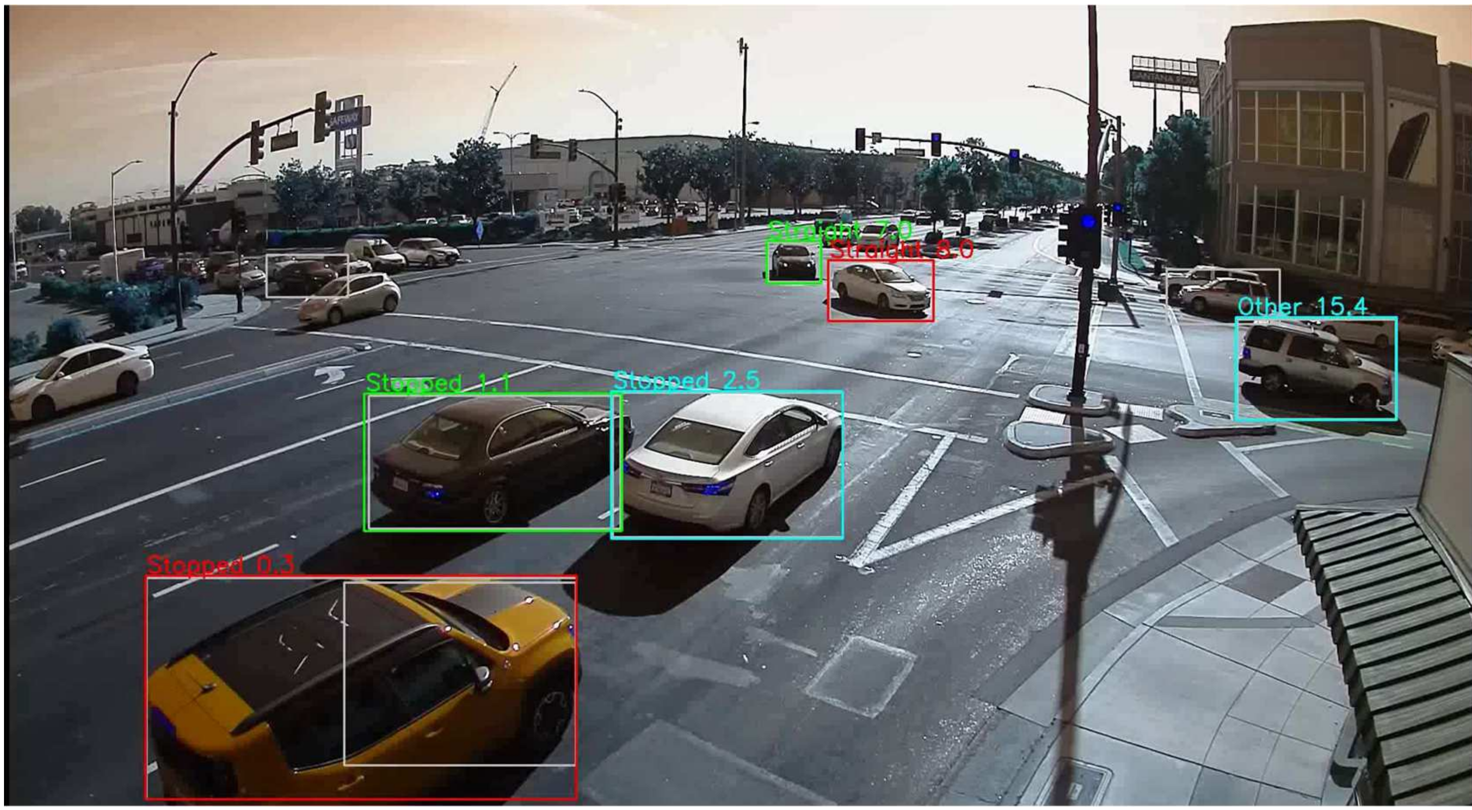
[Yi Wei, Nenghui Song, Lipeng Ke, Ming-Ching Chang, and Siwei Lyu, "Street Object Detection / Tracking for AI City Traffic Analysis", IEEE Smart World Congress, Honorable Mention Award, 2017]

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

2017 AI City Traffic Analysis Results

Stevens & Winchester 1:
Visualizing the moving direction & velocity (MPH) for each vehicle.



video length 1:10:20

Speed unit: mile per hour

Top-down view from Google Map

[Yi Wei, Nenghui Song, Lipeng Ke, Ming-Ching Chang, and Siwei Lyu, "Street Object Detection / Tracking for AI City Traffic Analysis", IEEE Smart World Congress, Honorable Mention Award, 2017]

UNIVERSITY AT ALBANY
State University of New York

From Seeing to Understanding

2018 NVidia AI City Challenge

Speed Estimation

4: CA locations
27: minutes of videos
56: teams downloading data
13: teams submitting results

Anomaly Detection

20+: IA locations
25: hours of all weather videos
53: teams downloading data
7: teams submitting results

Vehicle Re-Identification

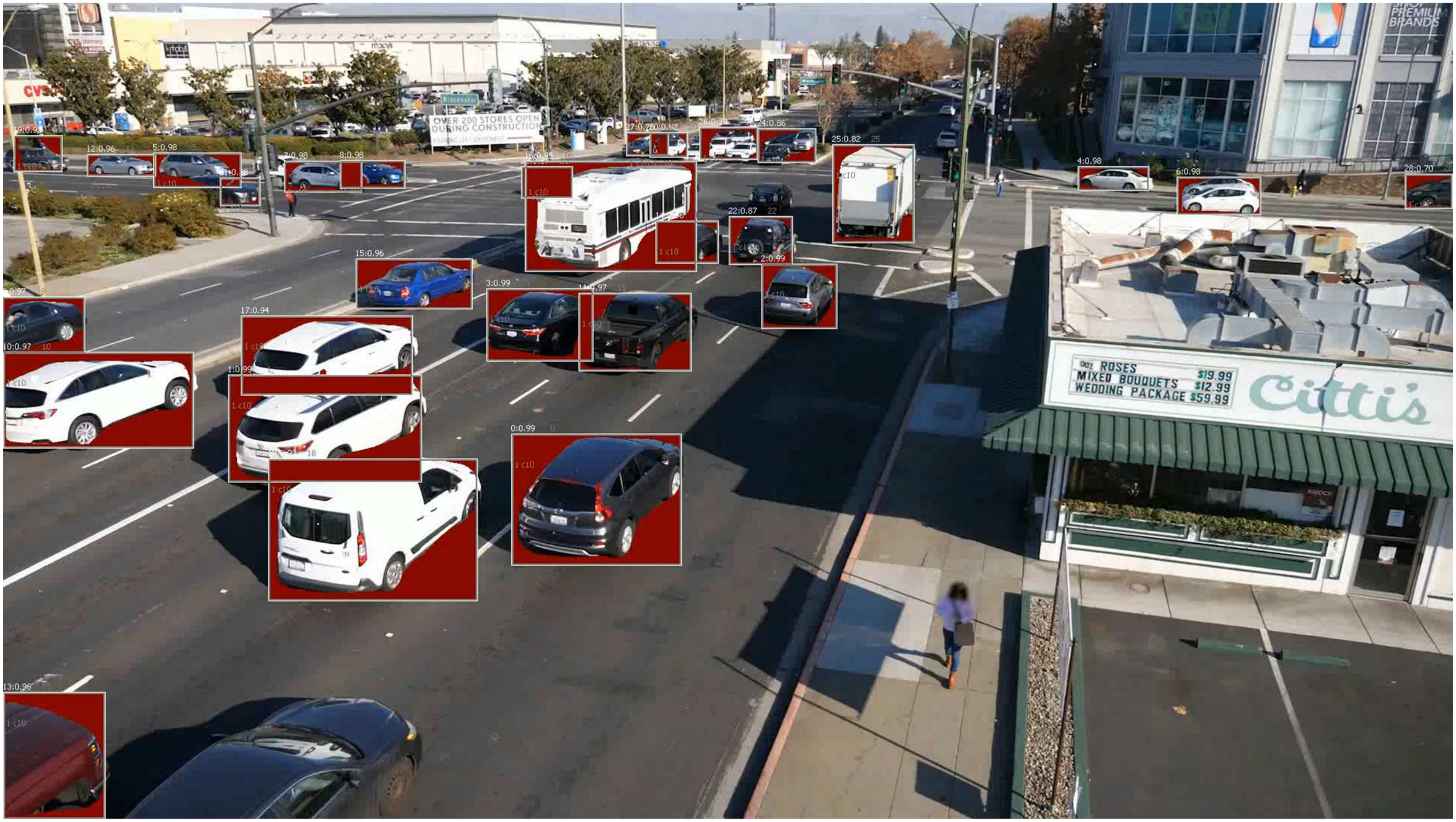
4: CA locations
15 hours of videos
61: teams downloading data
10: teams submitting results

[Milind Naphade, Ming-Ching Chang, Anuj Sharma, David C. Anastasiu, Vamsi Jagarlamudi, Pranamesh Chakraborty, Tingting Huang, Shuo Wang, Ming-Yu Liu, Rama Chellapa, Jenq-Neng Hwang, Siwei Lyu, "The NVIDIA AI City Challenge 2018", CVPR Workshop 2018]

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Mark R-CNN Detection and Tracking



[Ming-Ching Chang, Yi Wei, Nenghui Song, Siwei Lyu, "Video Analytics in Smart Transportation for the AIC'18 Challenge", IEEE CVPR Workshop on the NVIDIA AICity Challenge, 2018.]

Anomaly (Stopped Car) Detection

- Flow density estimation
- ROI Masking
- Feature extraction + anomaly classification

[Ming-Ching Chang, Yi Wei, Nenghui Song, Siwei Lyu, "Video Analytics in Smart Transportation for the AIC'18 Challenge", IEEE CVPR Workshop on the NVIDIA AICity Challenge, 2018.]

AVSS Traffic and Street Surveillance for Safety and Security (T4S) Workshop

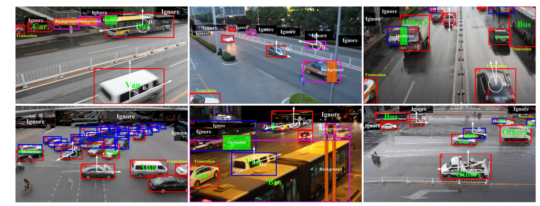
https://iwt4s.wordpress.com/

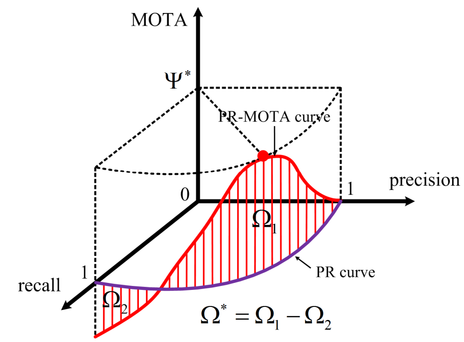
AVSS 2018

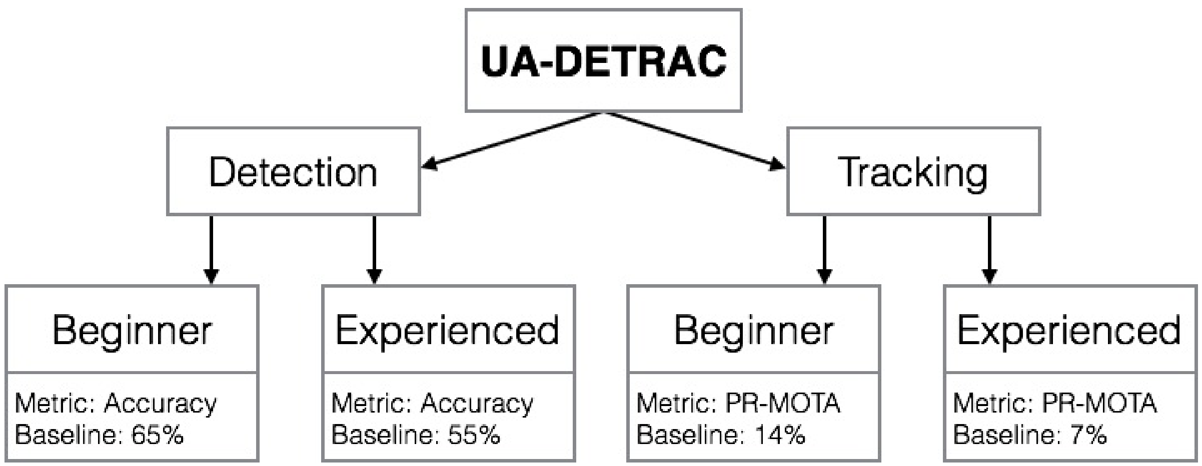
AVSS 2019

AVSS T4S Challenge

- Dataset – UA-DETRAC
- Evaluation:
  - Detection: average precision (AP)
  - Tracking: DETRAC PR-MOTA
- Challenges:







[S. Lyu, M.-C. Chang, D. Du, L. Wen, H. Qi, Y. Li, Y. Wei, L. Ke, T. Hu, M. D. Coco, P. Carcagni, *et al.*, "UA-DETRAC 2017: Report of AVSS2017 & IWT4S Challenge on Advanced Traffic Monitoring", AVSS 2017]

[S. Lyu, M.-C. Chang, D. Du, W. Li, Y. Wei, M. D. Coco, P. Carcagni, *et al.*, "UA-DETRAC 2018: Report of AVSS2018 & IWT4S Challenge on Advanced Traffic Monitoring", AVSS 2018]

Video Scenario Understanding

"Human Centric" Computing – understand situations, intentions, interactions
i.e., building a socially-aware agent

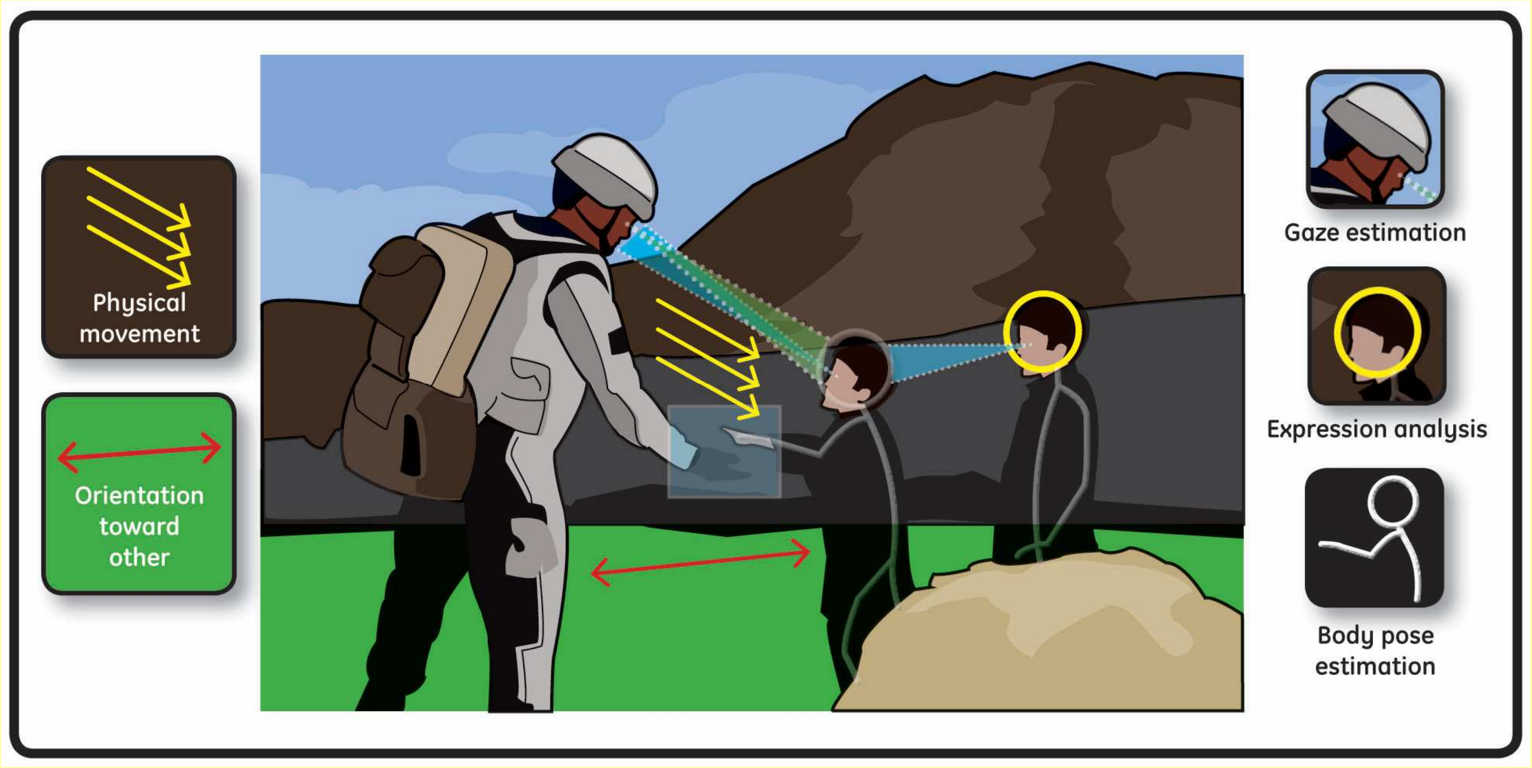


Image from GE DoD Sherlock System

Video Analytics

Reason about subjects, places, space and time.

Technology Pipeline:

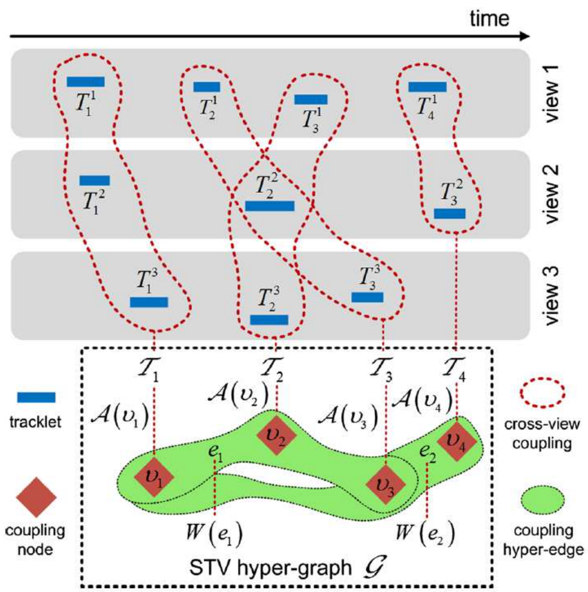
- **Detection** of Objects (Vehicles, Pedestrians, Faces, ...)
- **Tracking** – Kalman Filtering, Graph/Hypergraph based tracker
- **Mapping** – Camera Calibration, SfM & 3D Reconstruction
- **Analysis** – Rule-based or Learning based

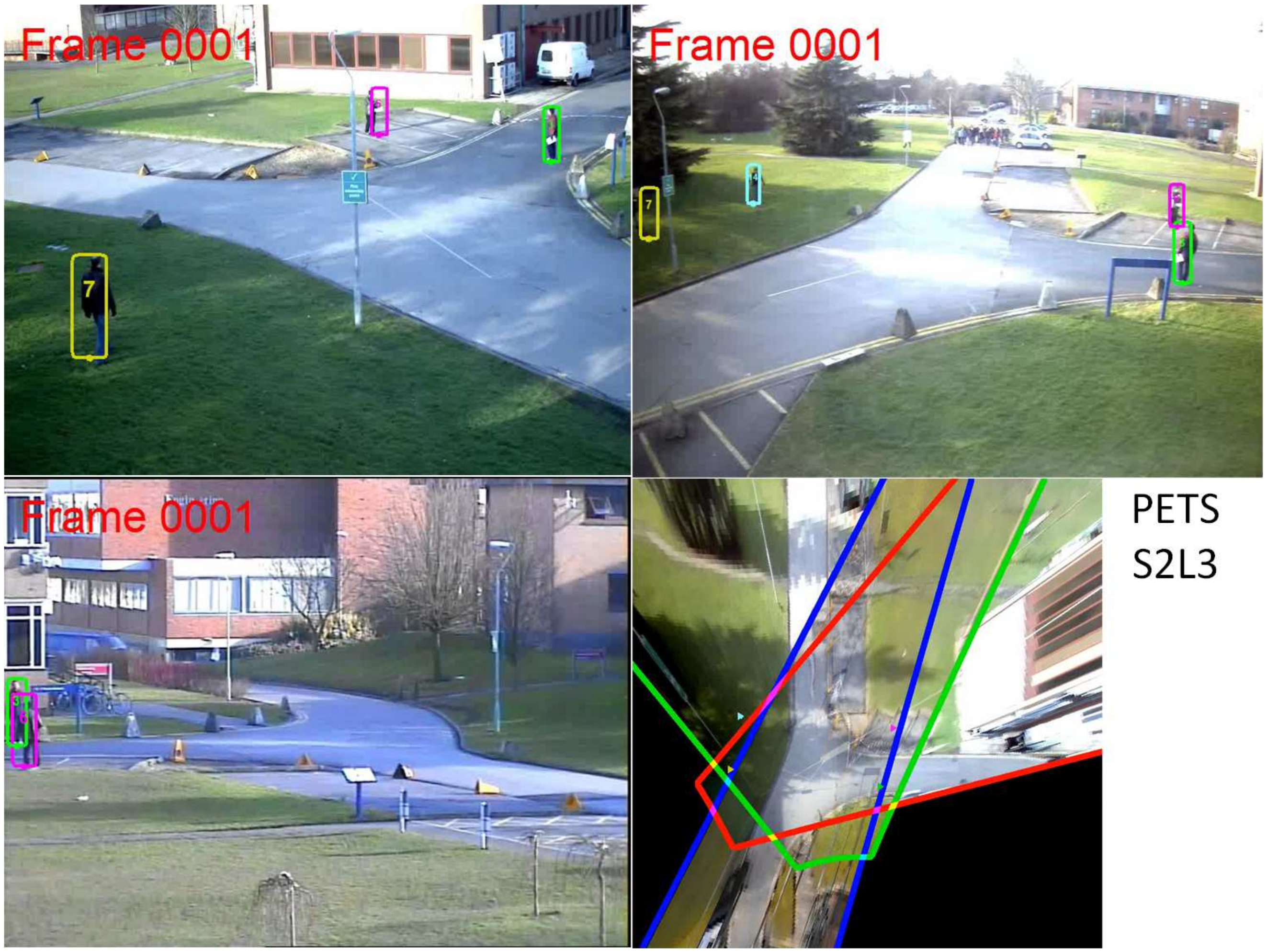
Topics: person, body poses, faces, activities.

Multi-View, Multi-Target Tracking

Input: noisy visual detections, known camera calibration

Space-Time-View Hypergraph



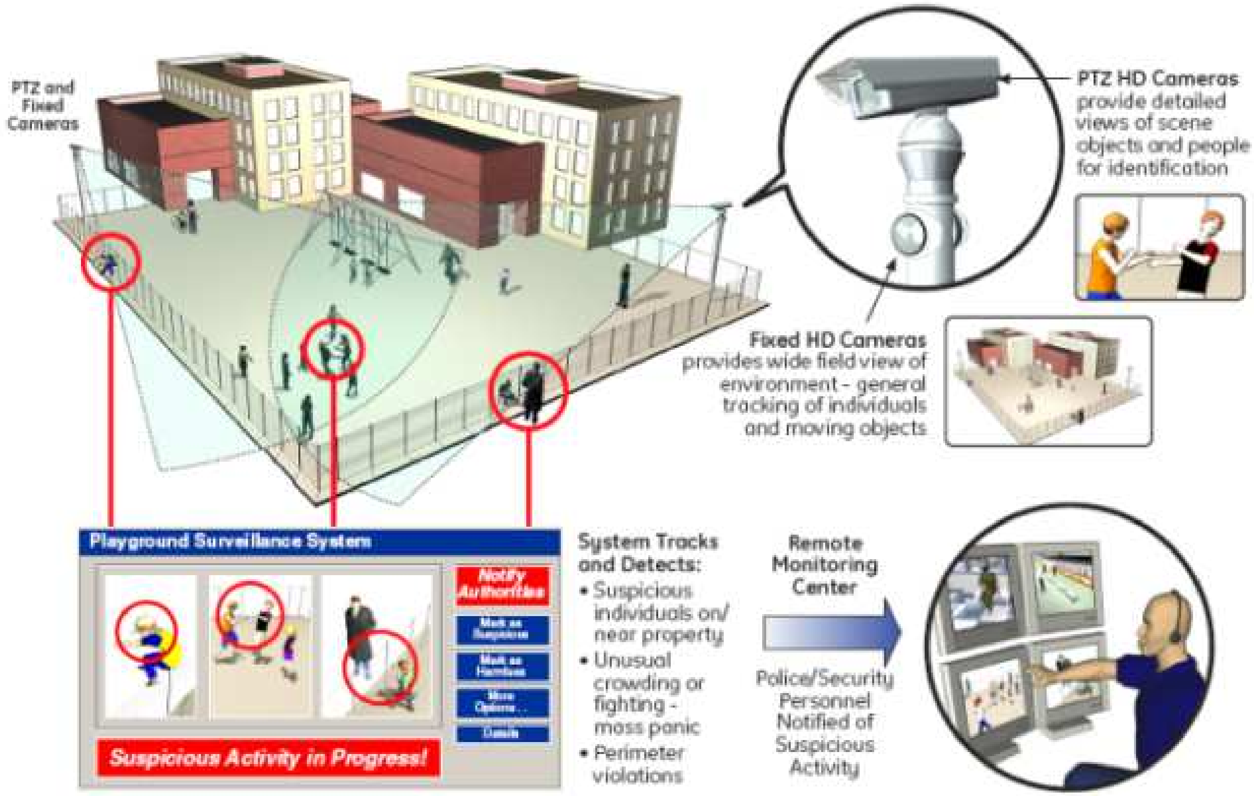


PETS S2L3

[Longyin Wen, Zhen Lei, Ming-Ching Chang, Honggang Qi, and Siwei Lyu, "Multi-Camera Multi-Target Tracking with Space-Time-View Hyper-graph", IJCV 2017]

NIJ Group Scenario Recognition

Continuous automated video-based behavior recognition in yards or public venues, where continuous law enforcement is desirable but infeasible.



Addressing security alerts involving **groups and crowds.**

[Ming-Ching Chang, Nils Krahnstoever, and Weina Ge, "Probabilistic Group-Level Motion Analysis and Scenario Recognition", ICCV 2011]

NIJ Mock Prison Riot

Scenario reasoning based on noisy tracking observations in real-time

Simulated scenarios. Data enacted by Lake Erie Correctional Officers.

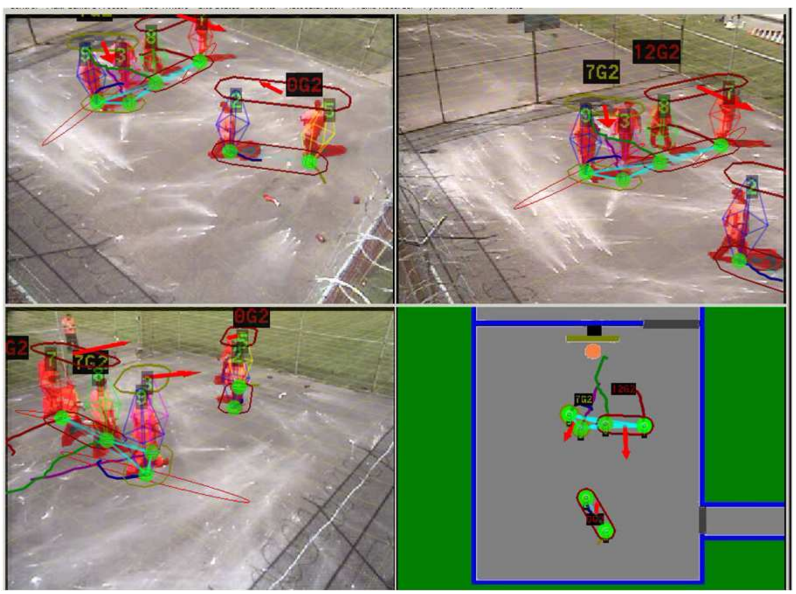
[Ming-Ching Chang, Nils Krahnstoever, and Weina Ge, "Probabilistic Group-Level Motion Analysis and Scenario Recognition", ICCV 2011]

UNIVERSITY AT ALBANY
State University of New York

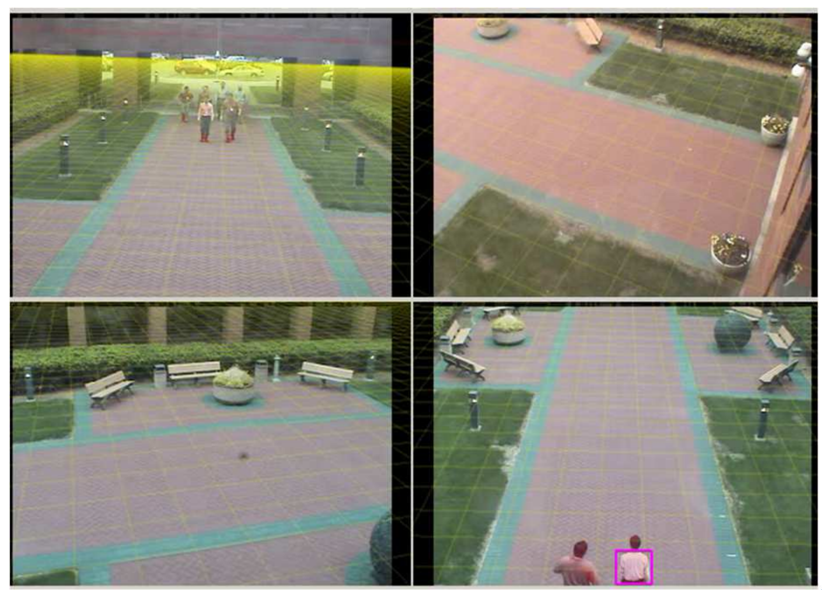
Group Formation / Dispersion

Object Hand-off

Flanking / Fighting



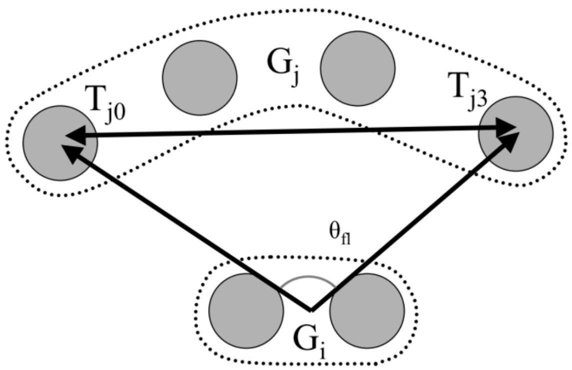
Chasing



[Ming-Ching Chang, Nils Krahnstoever, and Weina Ge, "Probabilistic Group-Level Motion Analysis and Scenario Recognition", ICCV 2011]

Detecting Flanking Maneuver

Aggressive group(s) surround victim group prior to an attack.



Criteria a Flanking Event:

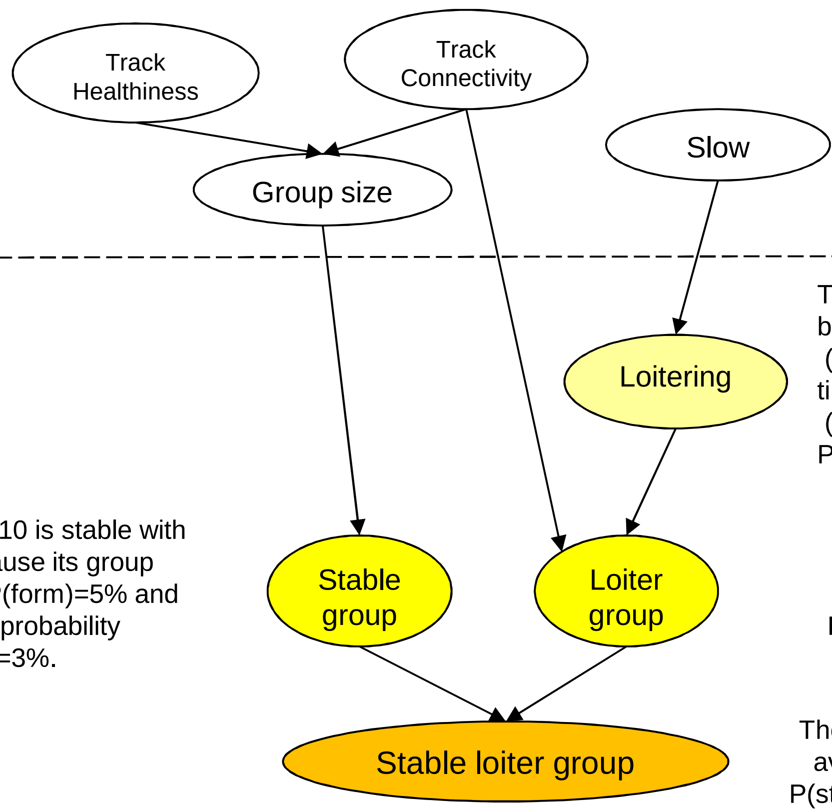
- (Attacking) individuals T_{j0} and T_{j3} are in same group, or are direct descendent of the same group, whose size is at least 2.
- Angle T_{j0}-G_i-T_{j3} exceeds a minimal angle θ_{fl} .
- Distance $d(T_{j0}, T_{j3})$ exceed the average of $d(T_{j0}, T_{j3})$ and $d(T_{j0}, T_{j3})$.

UNIVERSITY AT ALBANY State University of New York

[Ming-Ching Chang, Nils Krahnstoever, Sernam Lim, and Ting Yu, "Group Level Activity Recognition in Crowded Environments across Multiple Cameras", AMMCSS AVSS Workshop, ICCV 2010]

Event Explanation

Bayesian formulation intuitively allows explanation of a triggered event.



(Low-level detection with no explanation)

The loitering event is detected for target 10 with P(loiter)=71% because: (1) the target is currently moving slow with P(slow)=86%, (2) the target has been close to its current position at a point in time in the past within a window of 10.0 and 20.0 seconds ago, (3) the target was moving slow at that previous point in time with P(slow)=33%.

The grouping of target 10 is stable with P(stable)=92% because its group formation probability P(form)=5% and group dispersion probability P(disperse)=3%.

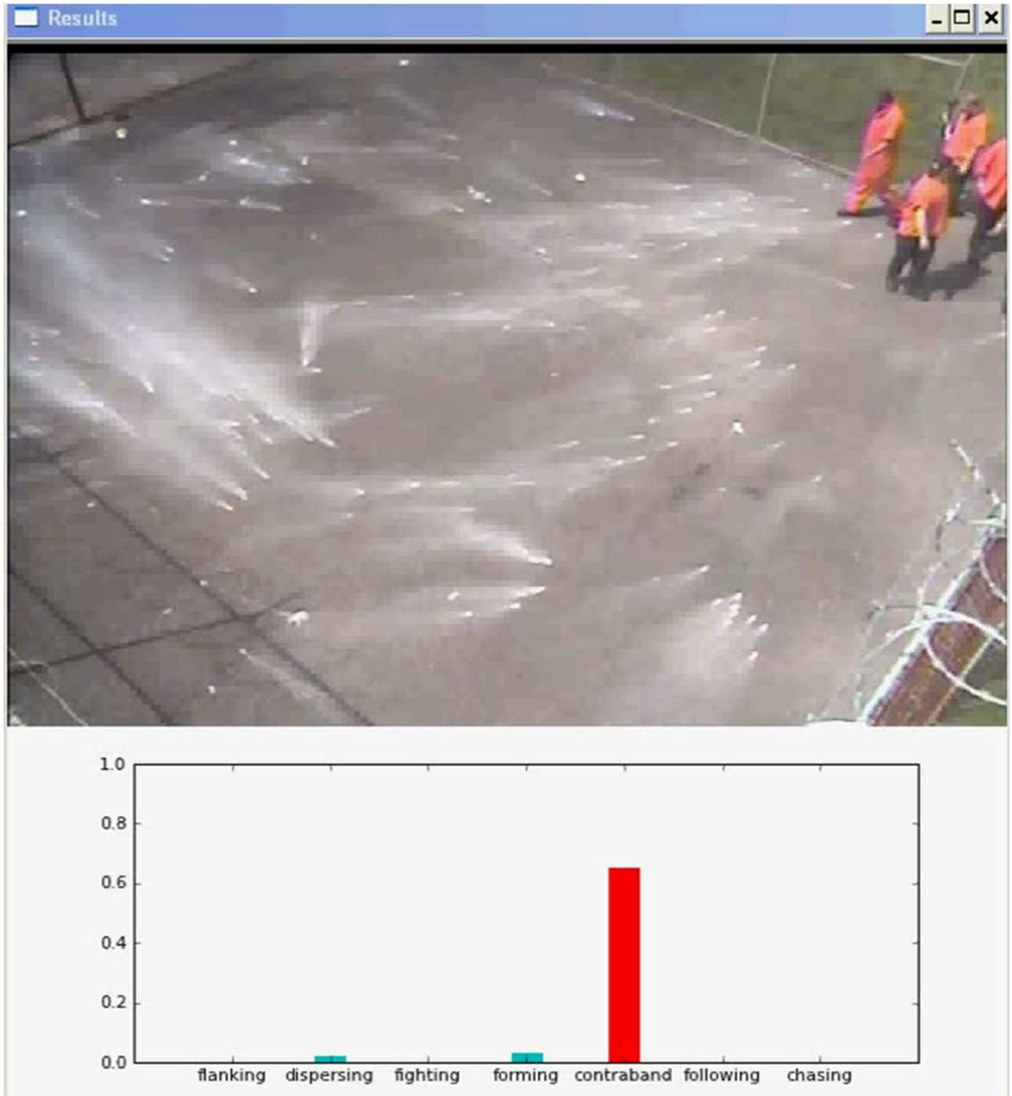
The loitering group event is detected for target 10 because its average loitering group probability from the soft grouping is P(group_loiter)=71%, which is greater than the threshold 60%.

The stable loitering group event is detected for target 10 because its average stable loitering group probability from the soft grouping is P(stable_loiter_group)=35%, which is greater than the threshold 30%.

All explanation texts are generated using semantic rule-based templates.

[Ming-Ching Chang, Nils Krahnstoever, and Weina Ge, "Probabilistic Group-Level Motion Analysis and Scenario Recognition", ICCV 2011]

Data-Driven Group Scenario Recognition



Different from the previous rule-based approach, here the recognition is achieved based on data-driven machine learning.

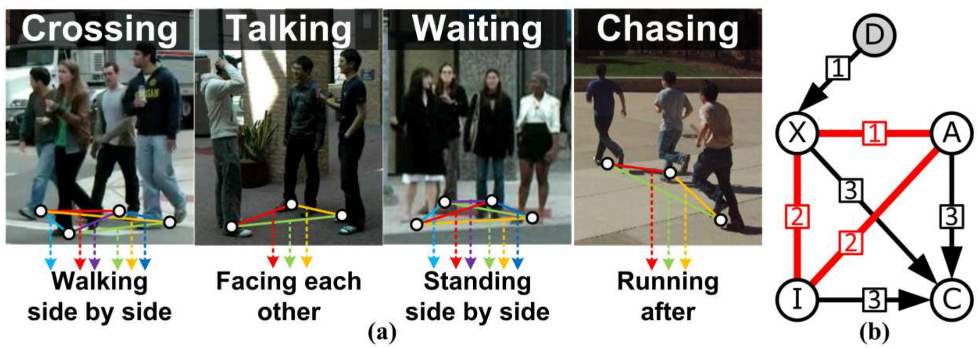
[Yimeng Zhang, Weina Ge, Ming-Ching Chang, and Xiaoming Liu, "Group-Level Context Learning for Event Recognition", WACV Best Student Paper, 2011]

[Yimeng Zhang, Xiaoming Liu, Ming-Ching Chang, Weina Ge, and Tsuhan Chen, "Spatio-Temporal Phrases for Activity Recognition", ECCV 2012]

UNIVERSITY AT ALBANY
State University of New York

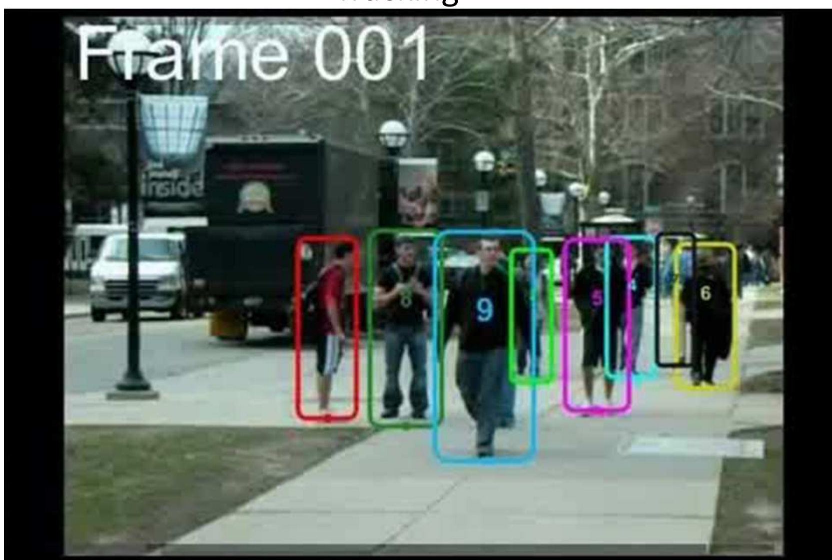
Group Behavior Recognition

Who did what at where and when –
Simultaneous tracking and group activity recognition



[Wenbo Li, Ming-Ching Chang, Siwei Lyu, "Who did What at Where and When: Simultaneous Multi-Person Tracking and Activity Recognition", arXiv 2018]

Tracking



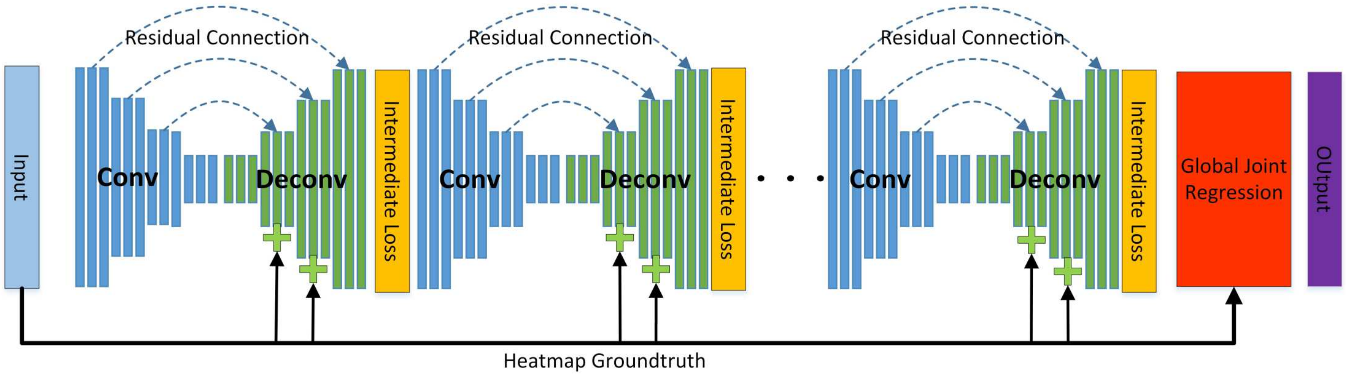
Group Activity Recognition

S standing
W walking
R running
AP approaching
LV leaving
PB passing by
FE facing each other
WS walking side-by-side
SR standing in a row
SS standing side-by-side
NA no-interaction
WO walking in oppo dir
WR walking after
RS running side-by-side

Michigan Collective Activity Datasets

Human Pose Estimation

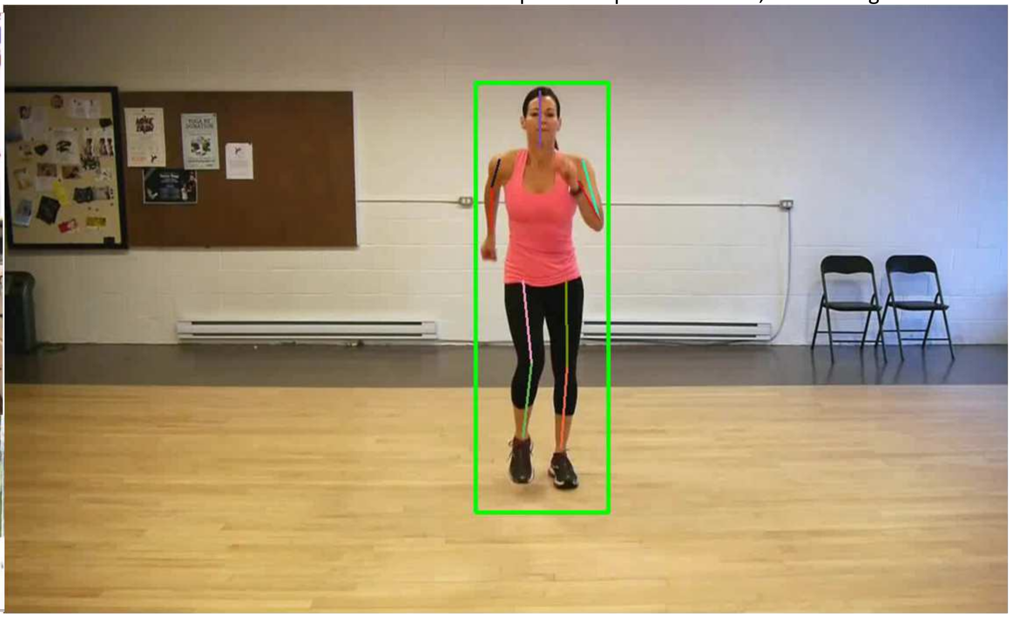
Hourglass modules with multi-scale supervision (for learned features) and global structural regression (to better handle scene clutters and multiple persons)



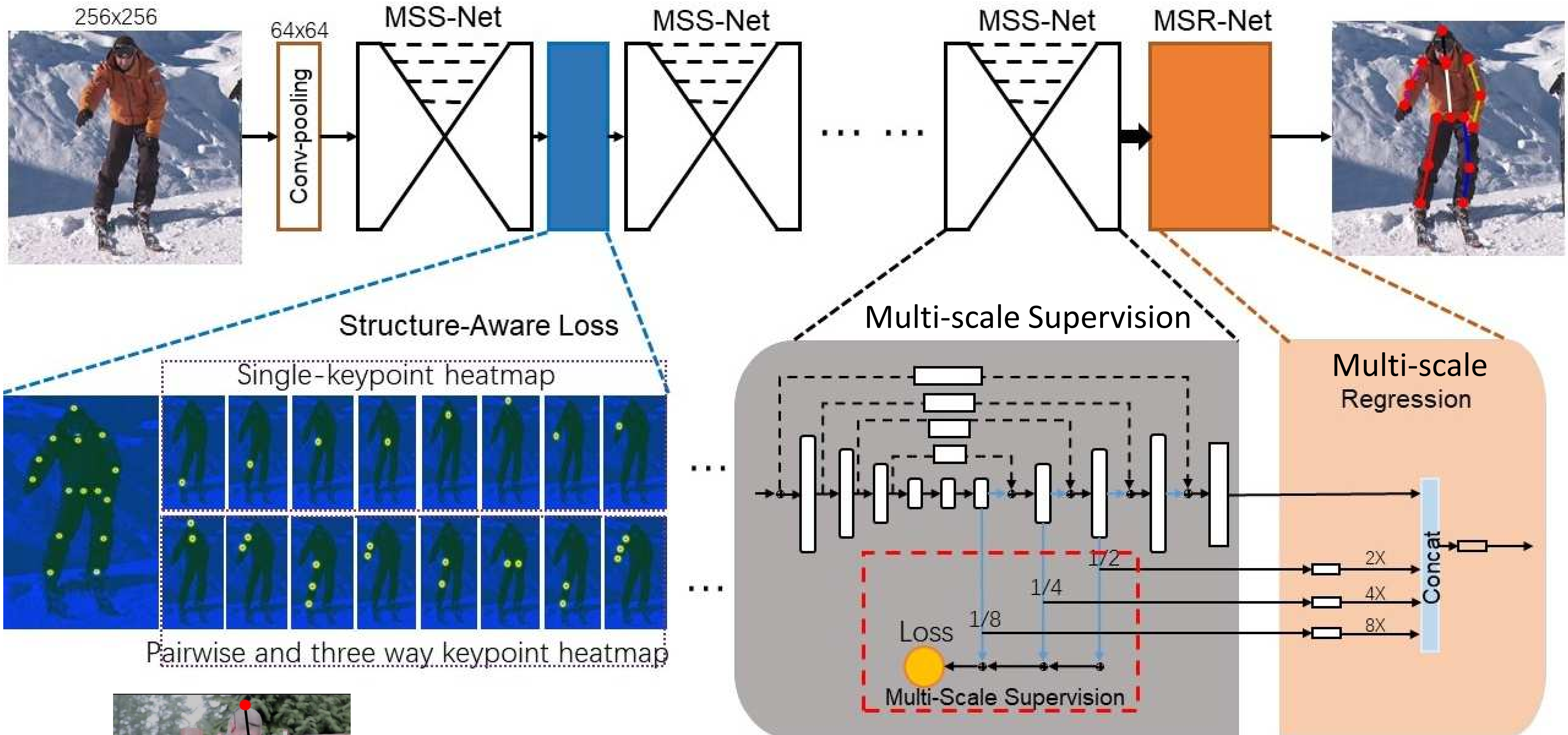
[Lipeng Ke, Ming-Ching Chang, Honggang Qi, Siwei Lyu, "Multi-scale Supervised Network for Human Pose Estimation", ICIP 2018]

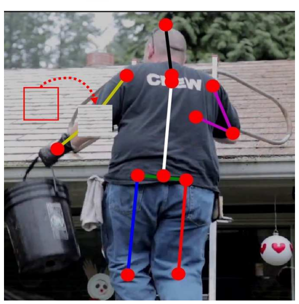
Example video: YouTube [Hi-lo Aerobic Combo Set #1]

Per-frame person & pose detection, no tracking.

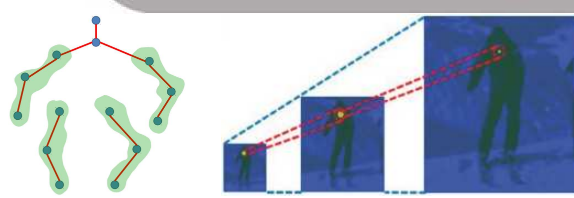


Multi-Scale Structure-Aware Network





Keypoint Masking
Training to improve
occlusion handling



[Lipeng Ke, Ming-Ching Chang, Honggang Qi, Siwei Lyu, "Multi-Scale Structure-Aware Network for Human Pose Estimation", ECCV 2018]

UNIVERSITY AT ALBANY State University of New York

FROM SEEING TO UNDERSTANDING

Multi-Scale Structure-Aware Network

This work was ranked top-1 in the MPII Human Pose Benchmark
http://human-pose.mpi-inf.mpg.de/ for the past 6 months. It is now top-2.

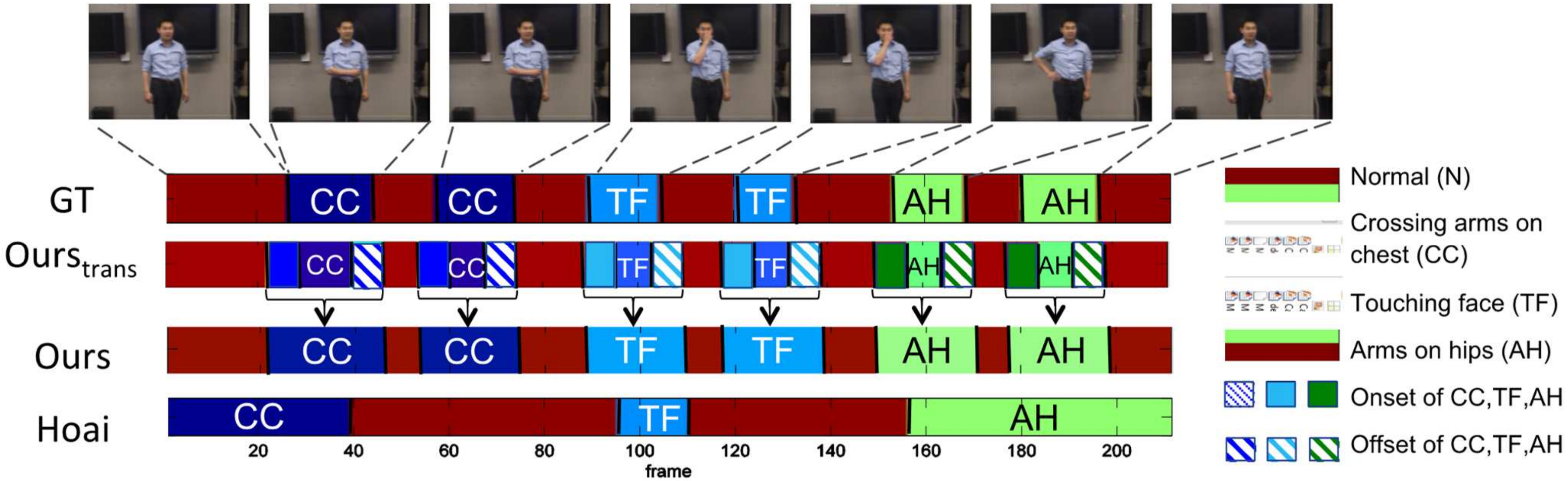
[Lipeng Ke, Ming-Ching Chang, Honggang Qi, Siwei Lyu, "Multi-Scale Structure-Aware Network for Human Pose Estimation", ECCV 2018]

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Action Recognition from Arm Tracking

- Jointly localize/segment events across time and recognize (classify) event types
- by modeling the transition between events
  - i.e. the onset and offset of 'touching face'

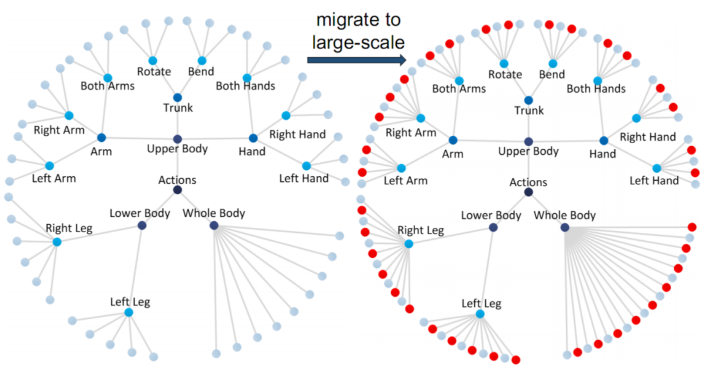


[Yelin Kim, Jixu Chen, Ming-Ching Chang, Xin Wang, Emily M. Provost, and Siwei Lyu, "Joint Event Localization and Classification of Human Action Videos with Event Transitions", FG 2015]

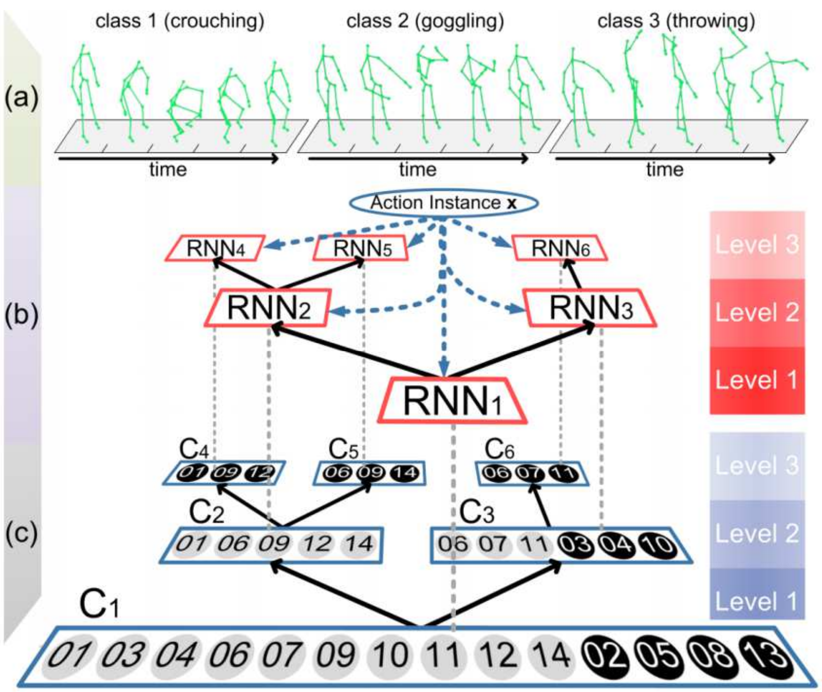
UNIVERSITY AT ALBANY
State University of New York

Action Recognition Using RNN Tree

[Wenbo Li, Longyin Wen, Ming-Ching Chang, Sernam Lim, and Siwei Lyu. "Adaptive RNN Tree for Large-Scale Human Action Recognition", ICCV 2017]



- Recurrent Neural Network on skeletal inputs
- hierarchy of categories for fine-grained recognition
- incremental learning



(a) Visualization of action instances from three action classes.
(b) A three-level RNN Tree (RNN-T) associated with the learned Action Category Hierarchy (ACH) in (c).
(c) Each circle represents an action class. Grey circles represent ambiguous classes, and black circles represent unambiguous ones. Action classes in the same box form one action category.

Facial Landmark Tracker

Facial landmark detection and tracking – for gaze and expression analysis

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Gaze Tracking Using Pan-Tilt-Zoom Cameras

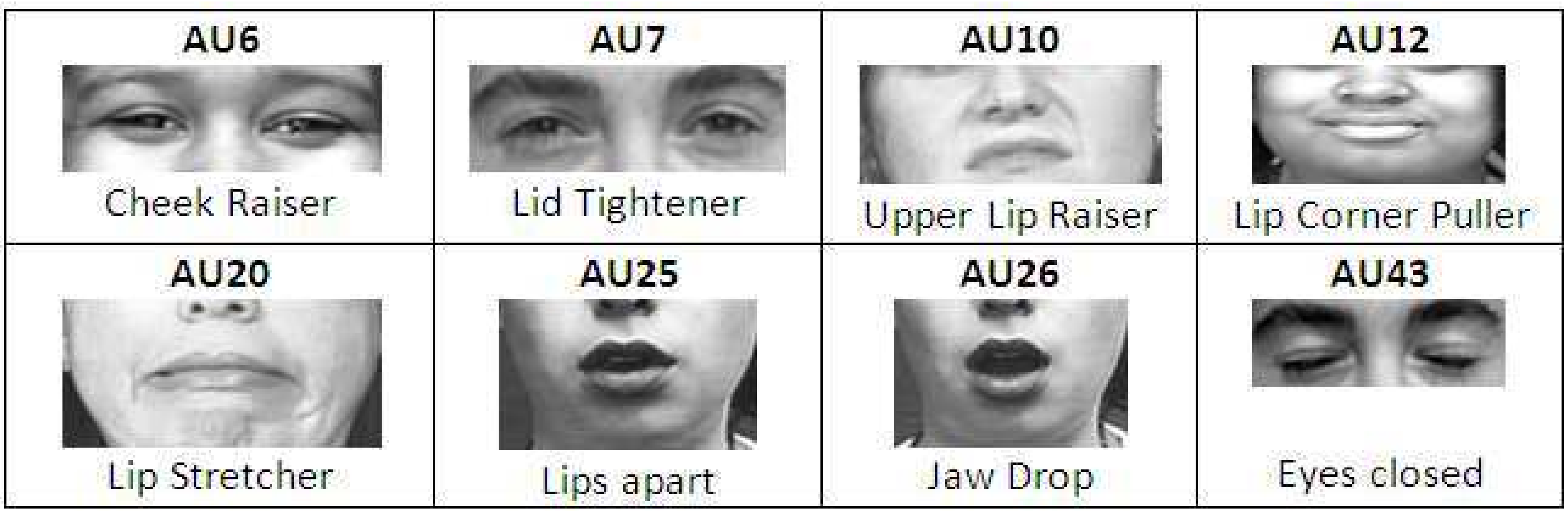
- Challenges:
  - Multi-sensors, multi-targets moving freely
  - Synchronization, noisy observations, resource optimization
- Fully automatic solution:
  - Particle filtering on person tracking and face detection
  - Body orientation implicitly inferred

[Nils Krahnstoever, Ming-Ching Chang, and Weina Ge, "Gaze and Body Pose Estimation from a Distance", AVSS Best Paper Award - Runner up, 2011]

Facial Analysis

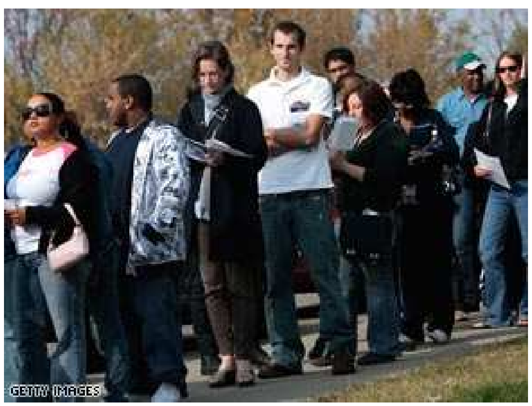
Expression Recognition

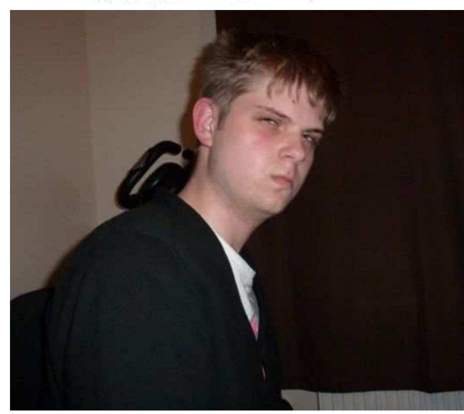
Facial Action Units



Body Language & Affect Analysis

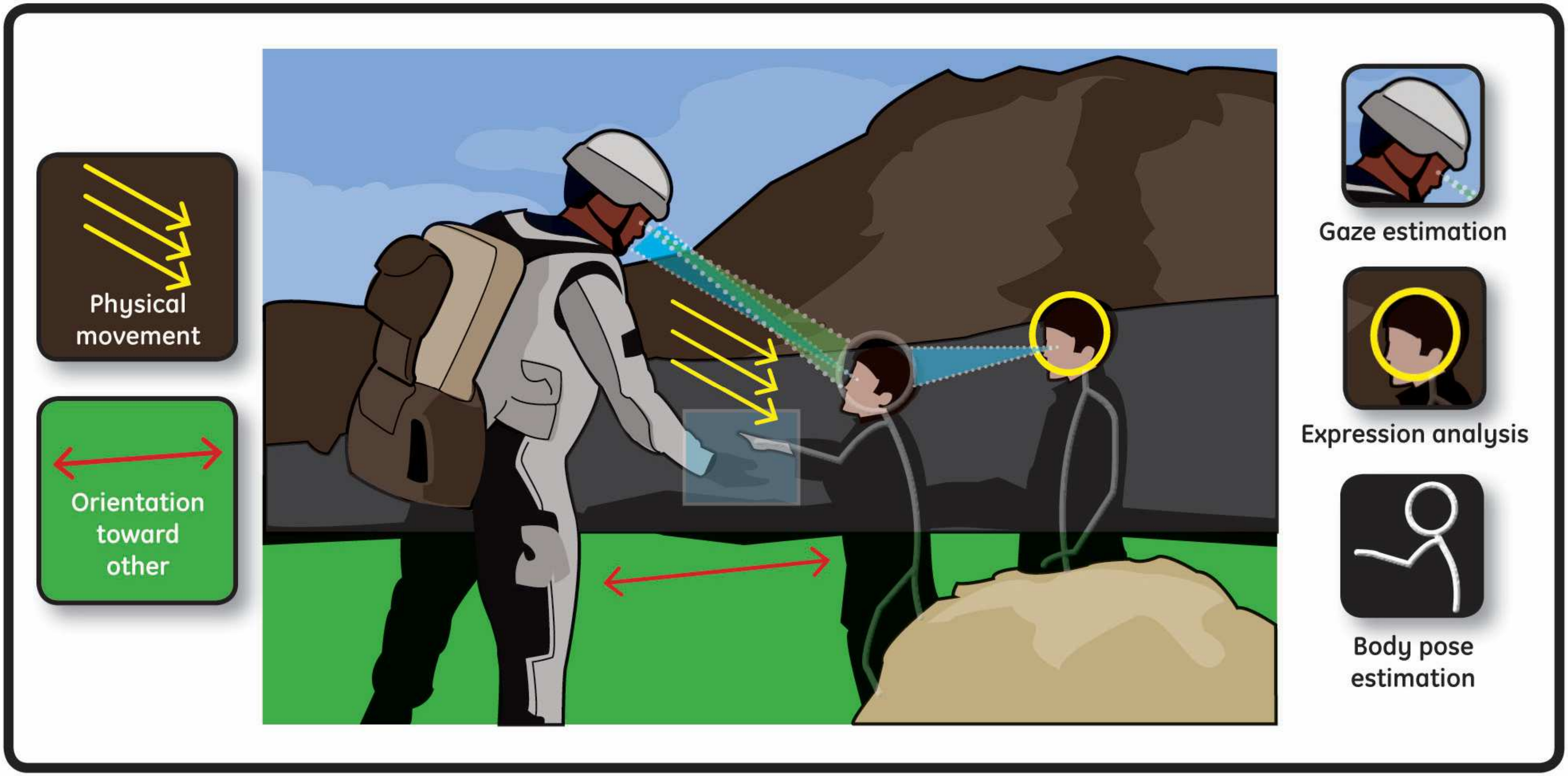
Risk Assessment via Visual Sensing





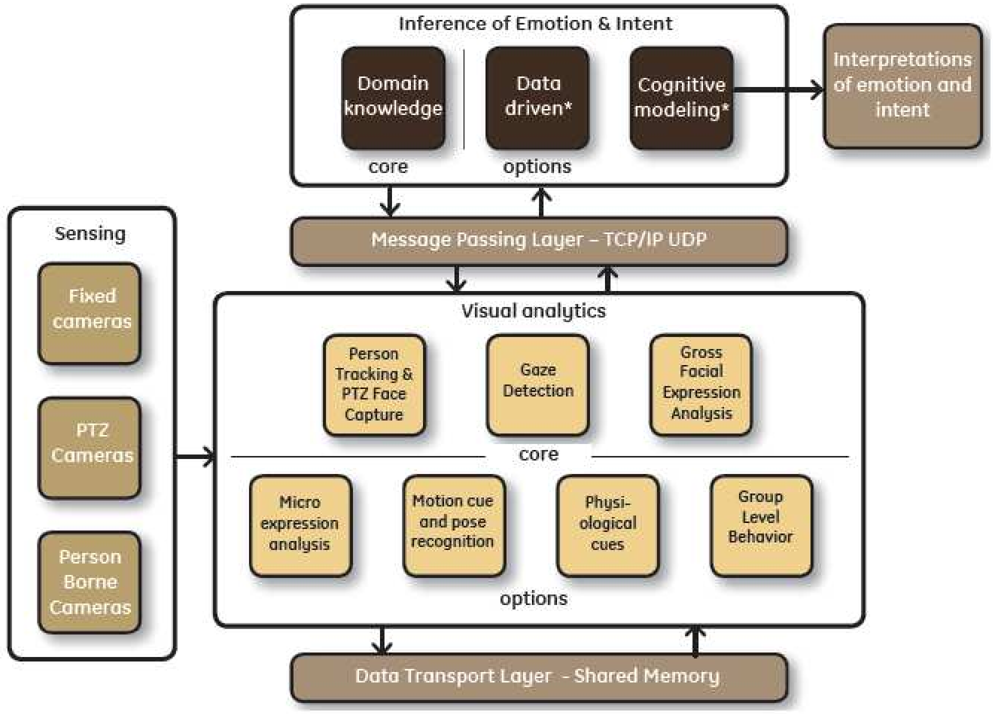
Social Interaction Analysis

GE Department of Defense (DoD) Sherlock System, 2013-2017

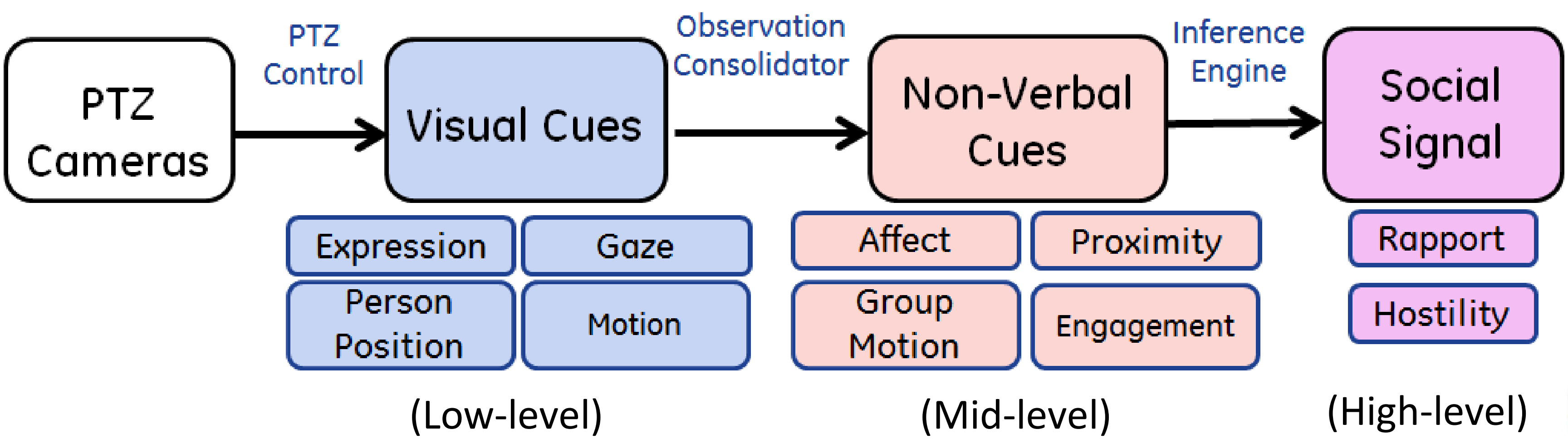


A Social Interaction Analysis Framework that maintains a number of sensors, video analytics and latent variable inference of internal emotional states and intentions of a group of individuals in a completely automated fashion.

System of Systems Architecture:

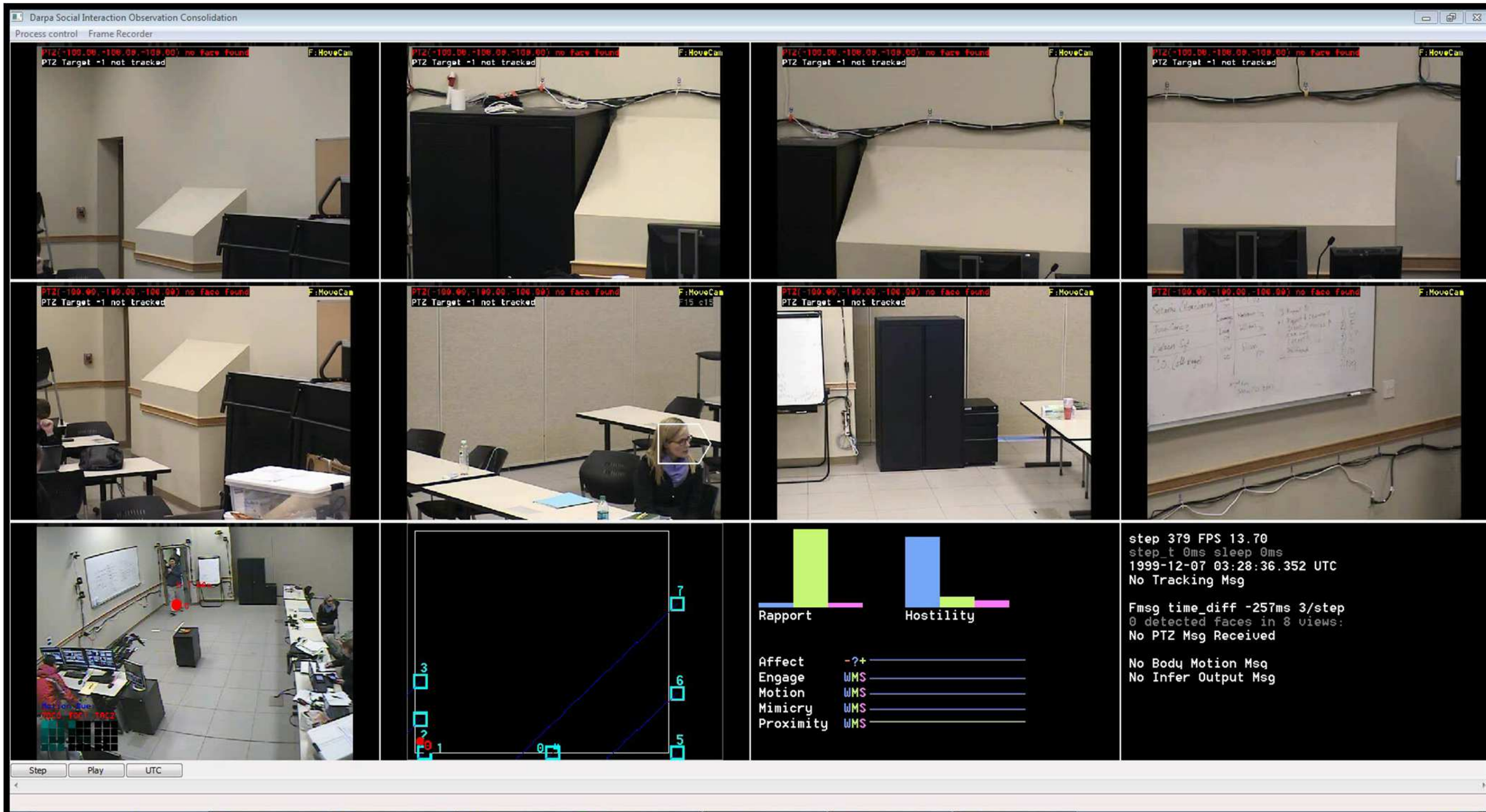


Process Flow: Hierarchical Cues for Inference



Live System at Army Site

Simulated military training session
- Automated driven PTZ based on person tracking
- Observation consolidator, social signal inference



[Peter Tu, Ming-Cing Chang, and Tao Gao, "Crowd Analytics via One Shot Learning and Agent Based Inference", GlobalSIP 2016]

Evaluation Related to Social Science Studies

- After Action Review -- comparison with social science labeling (groundtruth) and computer vision results
- Early-stage demo -- next-generation inference methods are under investigation (one-shot learning, agent-based simulation, RNN, …)

[Jixu Chen, Ming-Ching Chang, Tai-Peng Tian, Ting Yu, and Peter Tu, "Bridging Computer Vision and Social Science: A Multi-Camera Vision System for Social Interaction Training Analysis", ICIP 2015]

[Peter Tu, A. Logan-Terry, J. Chen, G. Rubin, Ming-Ching Chang, J. Hockett, Ting Yu, T.-P. Tian, "Cross-Culture Training Analysis via Social Science and Computer Vision Methods", AHFE 2015]

Industrial Video Analytics

Industrial Inspection

Healthcare

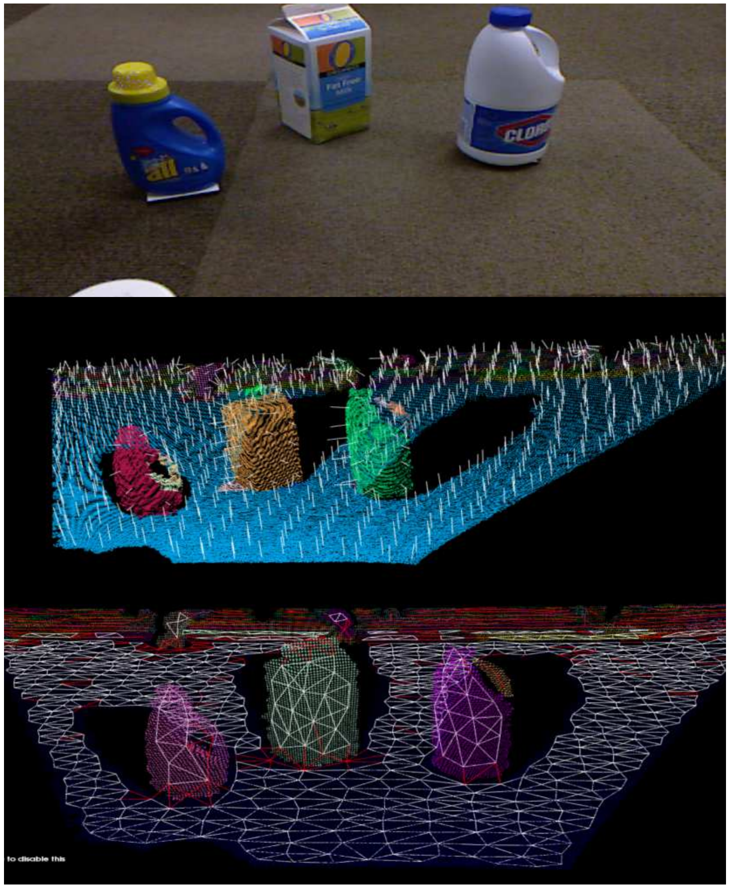
Situational Awareness

Site Management

Transportation

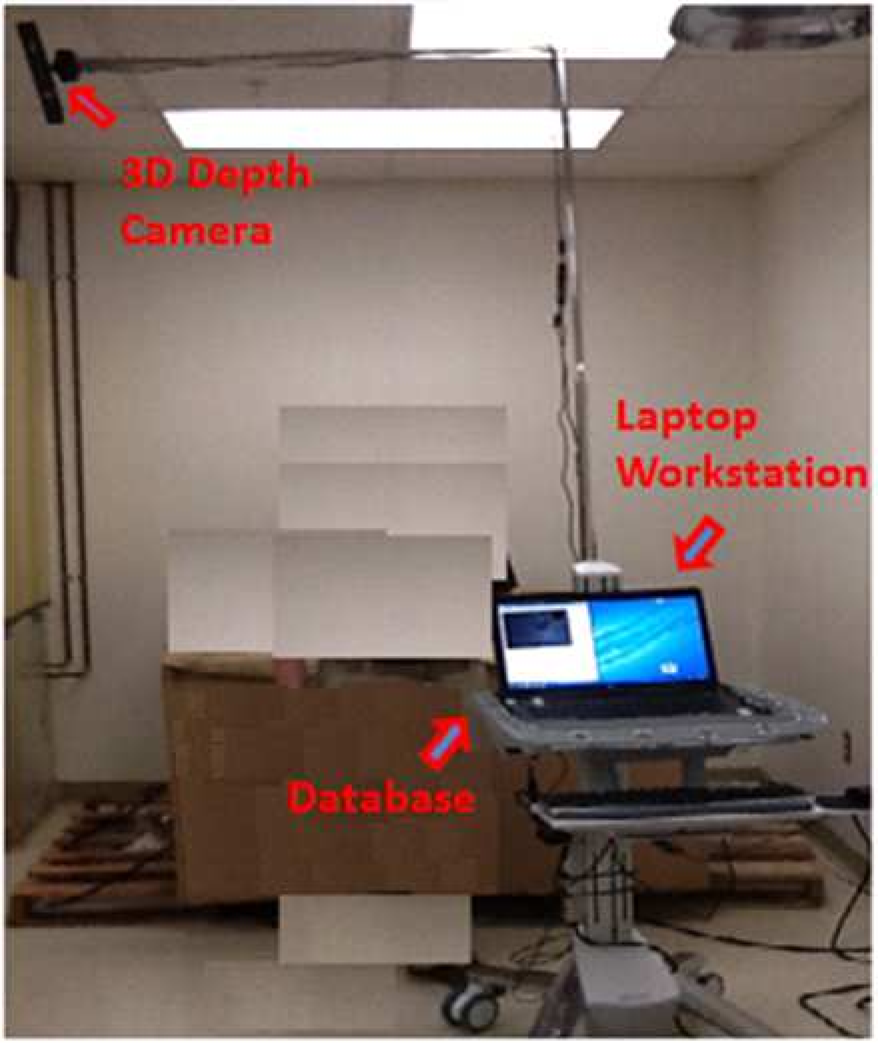
Robotics

- 2D/3D multi-modal (RGB, RGB+D, LiDAR) visual detection/tracking
- 3 DOF pose estimation
- robot navigation & control integration



Object segmentation and pose estimation for grasp planning

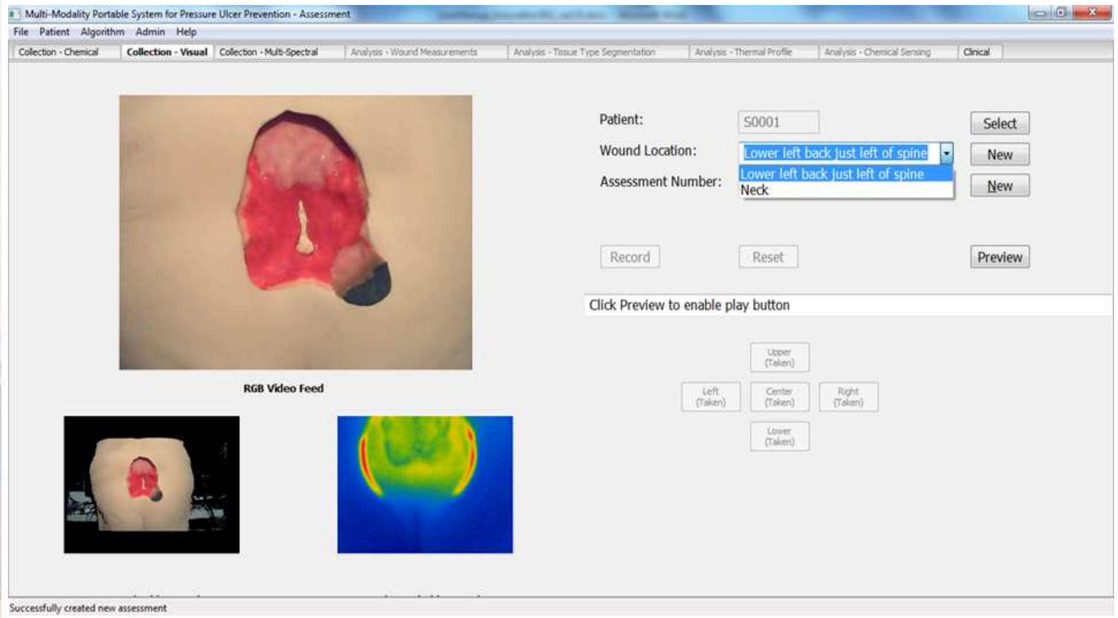
VA Pressure Ulcer Prevention System



U.S. Department of Veterans Affairs

[Ming-Ching Chang, Ting Yu, Jiajia Luo, Kun Duan, Peter Tu, Yang Zhao, Nandini Nagraj, Vrinda Rajiv, Michael Priebe, Elena Wood, and Max Stachura. "Multi-Modal Sensor System for Pressure Ulcer Wound Assessment and Care", IEEE Transactions on Industrial Informatics 2018]

VA Pressure Ulcer Assessment System

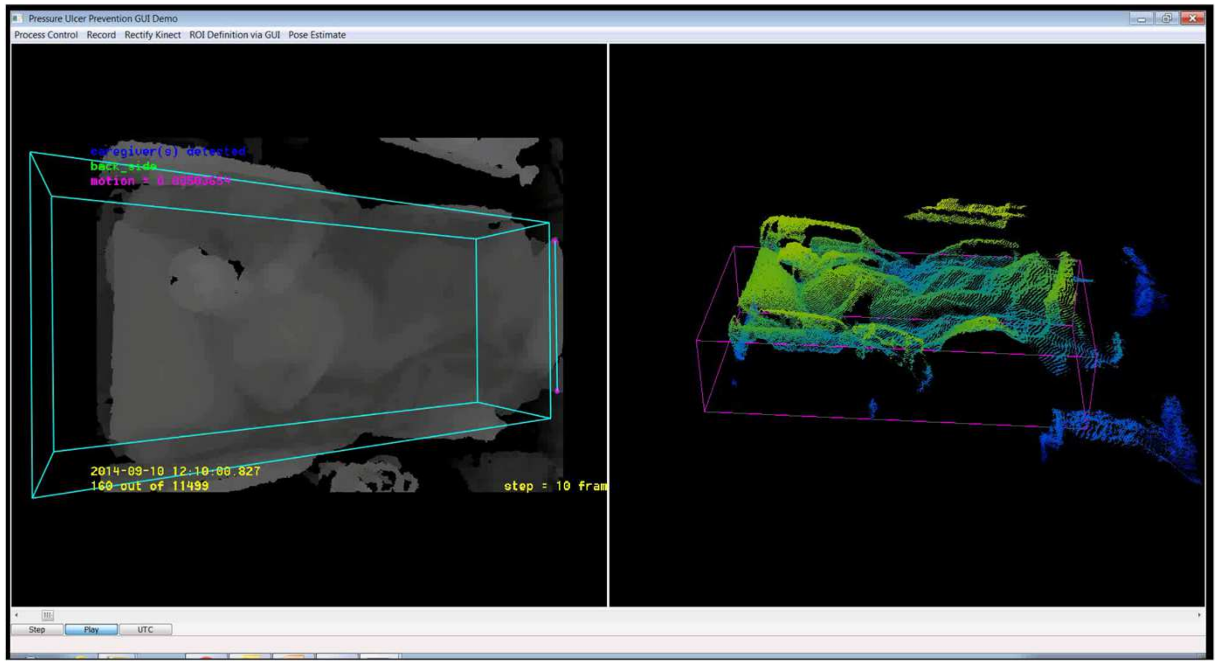


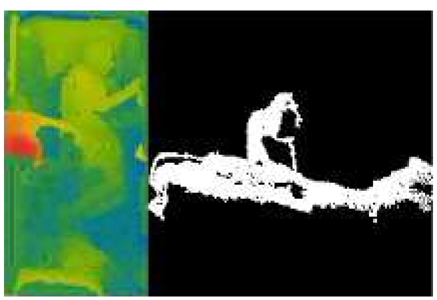
[Ming-Ching Chang, Ting Yu, Jiajia Luo, Kun Duan, Peter Tu, Yang Zhao, Nandini Nagraj, Vrinda Rajiv, Michael Priebe, Elena Wood, and Max Stachura. "Multi-Modal Sensor System for Pressure Ulcer Wound Assessment and Care", IEEE Transactions on Industrial Informatics 2018]

Pose Monitoring for Pressure Ulcer Prevention

Raw depth video with pose recognition result

3D point cloud after coordinate rectification

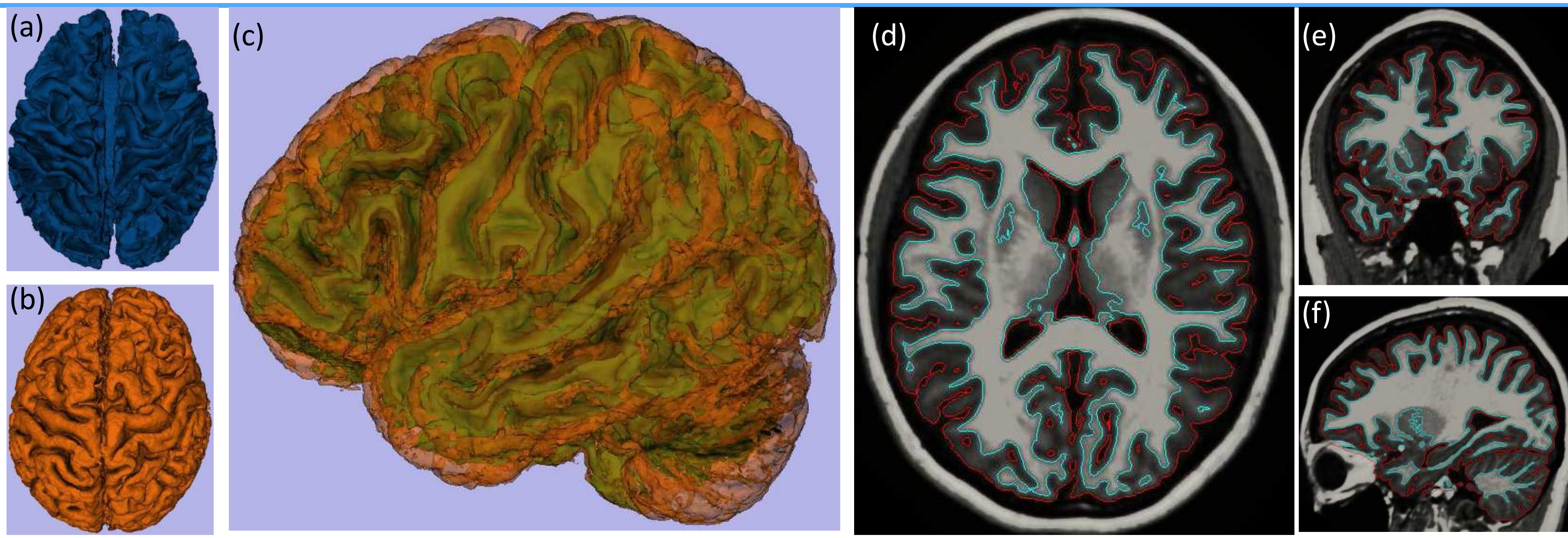


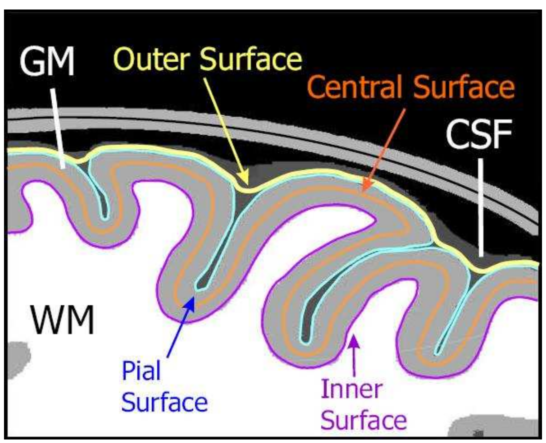


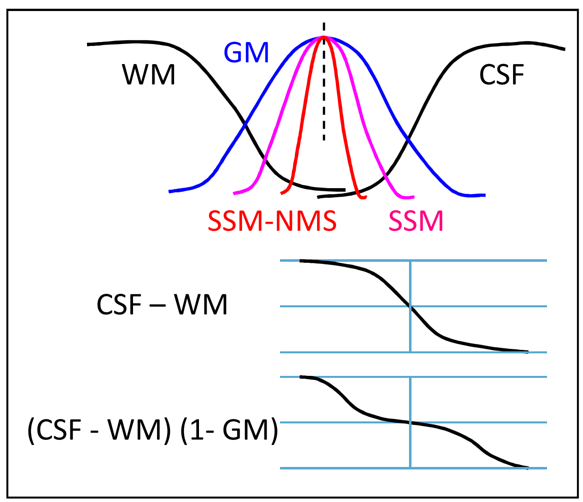
Method: (1) rectification and coordinate transform, (2) aggregate motion signals over time as a motion history image (MHI) for feature extraction and pose classification.

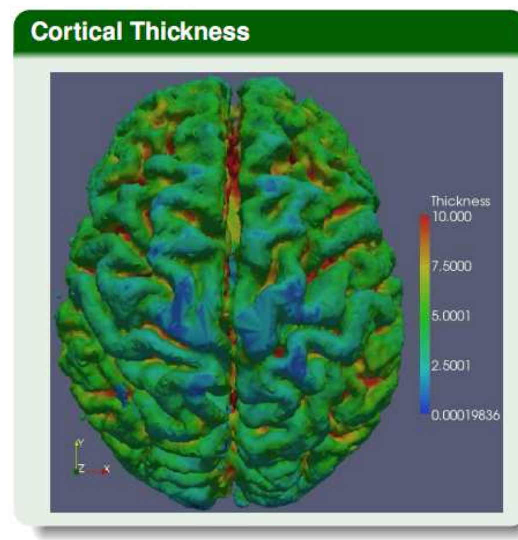
[Ming-Ching Chang, Ting Yu, Kun Duan, Jiajia Luo, Peter Tu, Michael Priebe, Elena Wood, and Max Stachura, "In-Bed Patient Motion and Pose Analysis Using Depth Videos for Pressure Ulcer Prevention", ICIP, 2017]

MRI Neural Image Analysis









[Ming-Ching Chang and Xiaodong Tao, "Subvoxel Segmentation and Representation of Brain Cortex using Fuzzy Clustering and Gradient Vector Diffusion", SPIE Medical Imaging 2011]

Improved DCE-MRI of Prostate at 3T with B_1 Correction

Bloch-Siegert based B_1 inhomogeneity correction of Variable Flip Angle T_1 mapping

Apply same B_1 correction to DCE SPGR signal intensity to [Gd] concentration conversion for pharmacokinetic (PK) analysis

- Tested on 13 patents; Yielded improved T_1 quantification
- Improves PK maps and lesion conspicuity by incorporating B_1 correction

[Ming-Ching Chang, Sandeep N. Gupta, Laura I. Sacolick, Clare Tempany-Afhdal, Fiona Fenessey, Ehud Schmidt, "Improved T1 Mapping and DCE MRI Pharmokinetic Quantification for Prostate at 3T by Incorporating B1 Inhomogeneity Correction", ISMRM 2013]

Collaboration with **Brigham and Women's Hospital**, Boston; Research supported by NIHU01CA151261

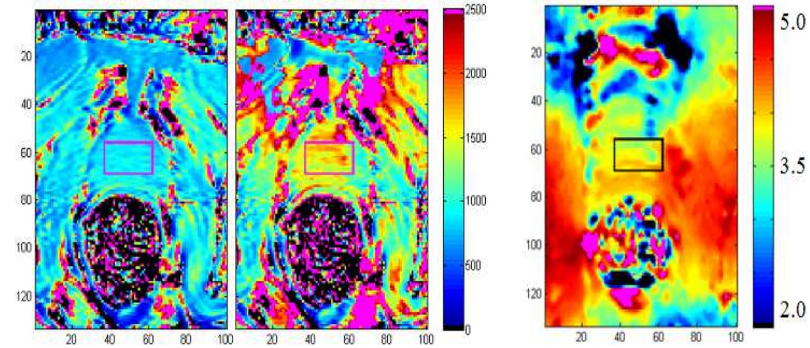
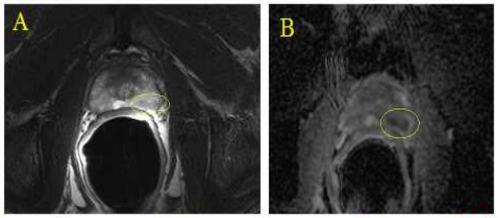
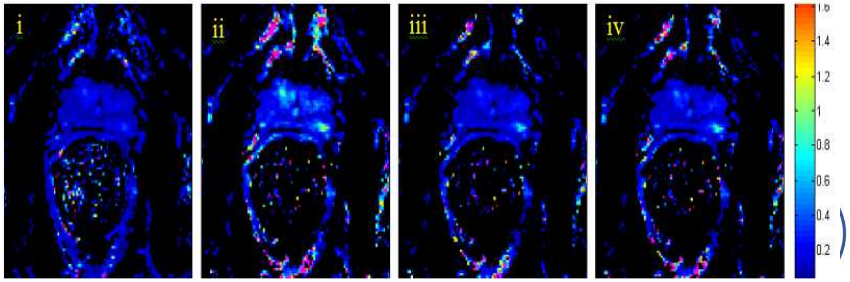


Fig.1: T_1 map (ms) before (left) and after (middle) B_1 correction, and corresponding B_1 map (µT) (right). Average ROI $T_{1before}$ = 895.80 ms, T_{1after} = 1540.59 ms.

	TR=15ms (13 cases)	
T_1 (ms)	W/o correction	B_1 correction
mean	977.87	1261.89
std/mean	0.12	0.15

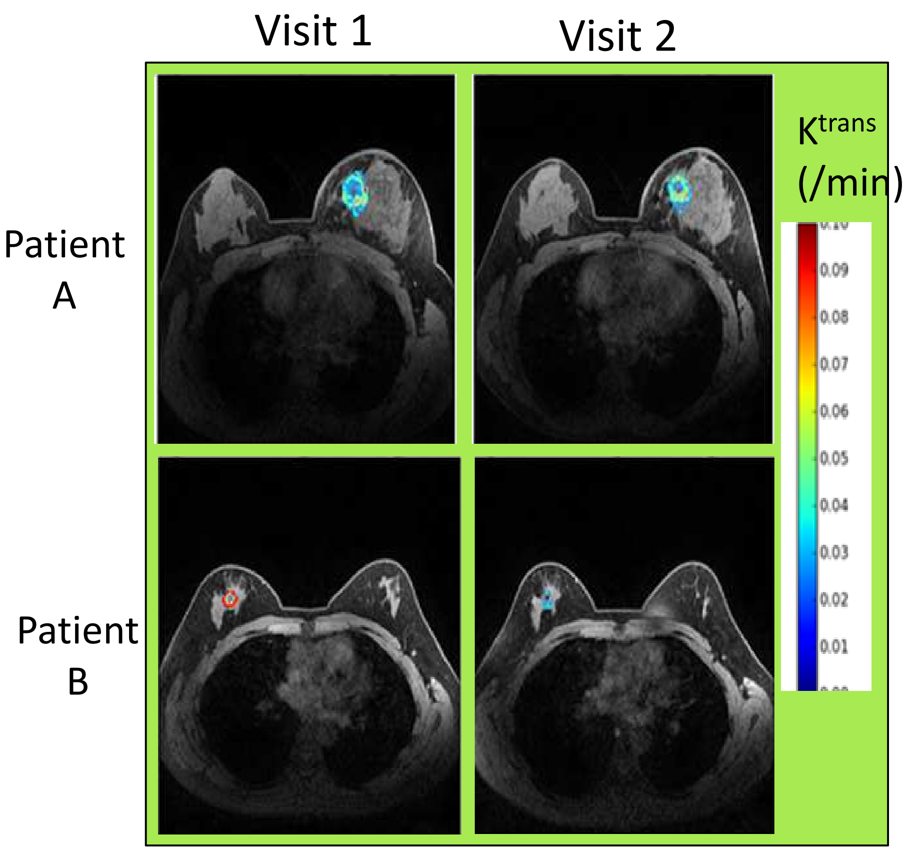
Fig.2: T_2w image (A); ADC map (B); K^{trans} maps for 4 methods (i)-(iv). Yellow highlighted region denotes confirmed tumor.





NIH Breast DCE Challenge

- 10 cases pre and post treatment.
- **Objective**: Determine tumor response to treatment using K^{trans} permeability
- Phase 1: assess treatment response in first 2 subjects.
- Phase 2 : 8 more subjects being analyzed.



- External AIF provided with the challenge data
- 2-compartment PK Model

$$C_t(t) = K^{trans}[C_p(t) \otimes e^{(-k_{ep}*t)}]$$

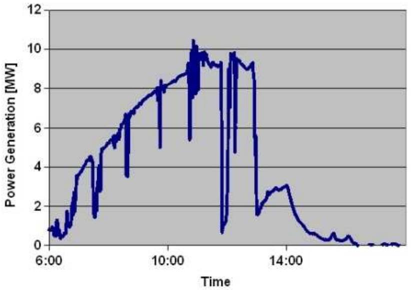
Patient A (Non responder) Ktrans (/min)		
Visit 1	Visit 2	Delta
0.031 (0.045)	0.032 (0.057)	-3.23%

Patient B (Complete Responder) Ktrans (/min)		
Visit 1	Visit 2	Delta
0.070 (0.058)	0.017 (0.014)	75.71%

[Wei Huang and Xin Li and Yiyi Chen and Xia Li and Ming-Ching Chang et al., “Variations of Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Evaluation of Breast Cancer Therapy Response: A Multicenter Data Analysis Challenge”, Translational Oncology, 2014]

Solar Forecaster

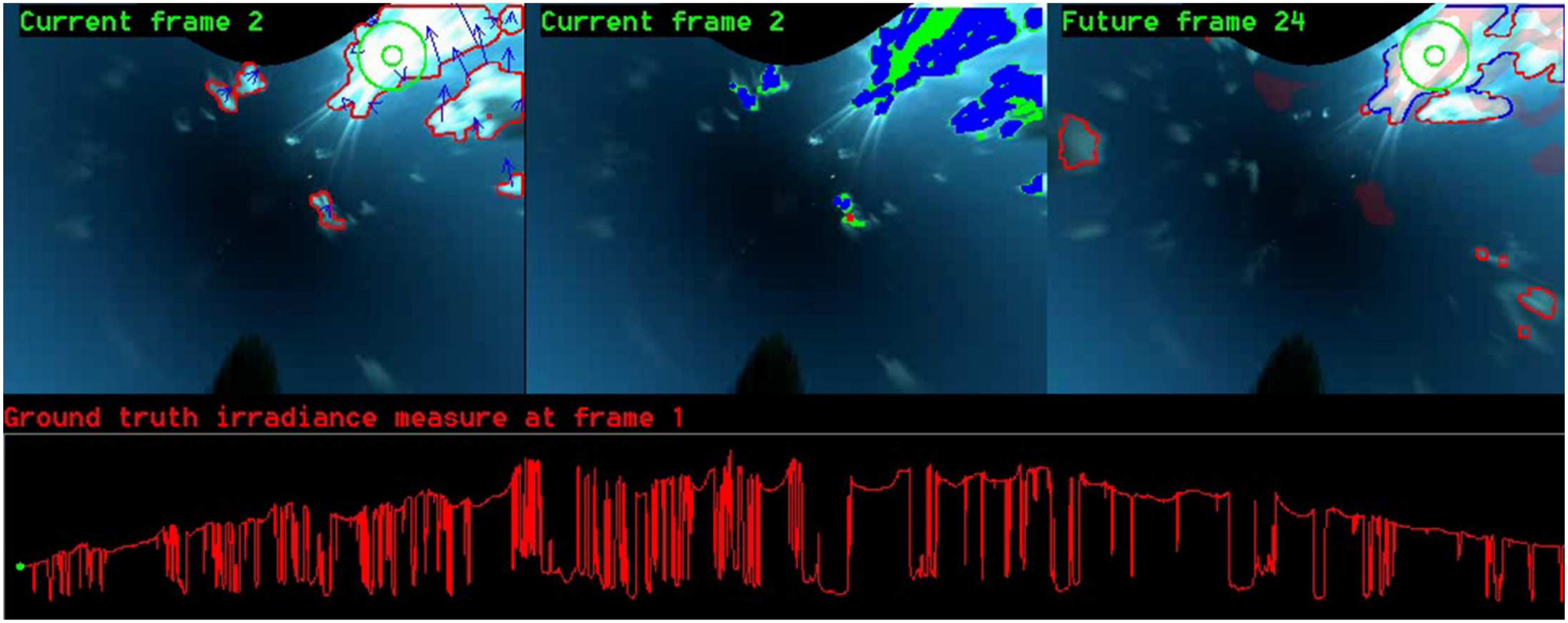
Clouds passing over MW PV plants



Steep ramp rates in plant power

Grid integration issues

[Ming-Ching Chang, Yi Yao, Li Guan, Yan Tong, and Peter Tu, "Cloud Tracking for Solar Irradiance Prediction", ICIP 2017]

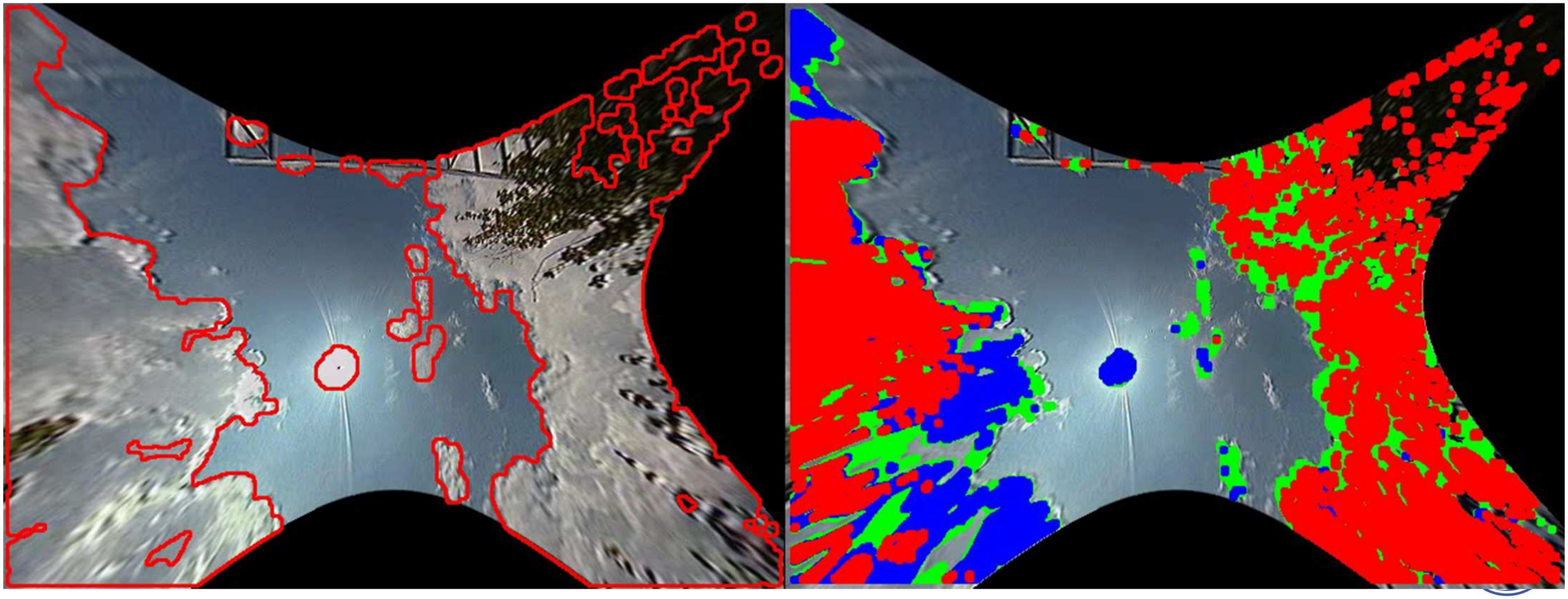


Sky Imager for Cloud Tracking

Steps and modules:

- Sun tracking & exposure control
- Fisheye rectification
- Cloud segmentation, classification and tracking
- Sun occlusion prediction

[Ming-Ching Chang, Yi Yao, Li Guan, Yan Tong, and Peter Tu, "Cloud Tracking for Solar Irradiance Prediction", ICIP 2017]



Automatic PTZ View Separation & Merging

PTZ in pivoting mode

[Andrew Pulver, Ming-Ching Chang, and Siwei Lyu, "Shot Segmentation and Grouping for PTZ Camera Videos", ASIA 2017]

View-split with background stabilization
Remove all transitional frames

The 5 pre-set PTZ pivot view are automatically recognized and separated.

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

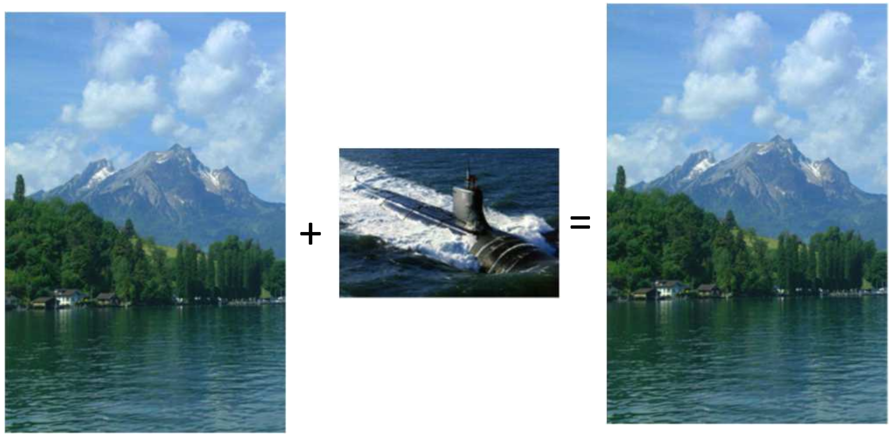
Railcar Text ID Detection & Recognition

For:
- Railcar management
- Counting
- Type classification
- Tracking, identification

Digital Media Forensics

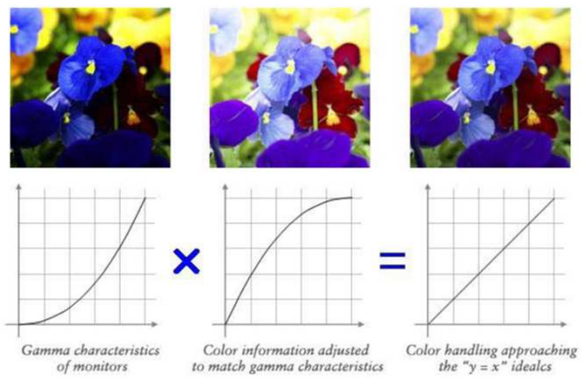
Image Forensic

CG or photo



detect image steganography (concealing info)

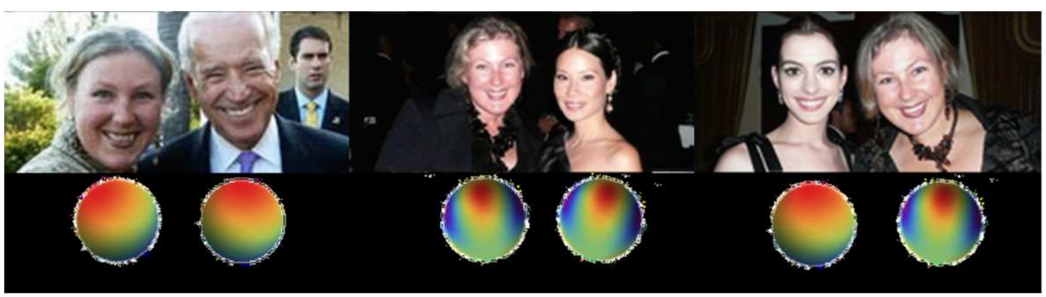
detect lens authentication



detect contrast enhancement

detect region duplication

detect face composite

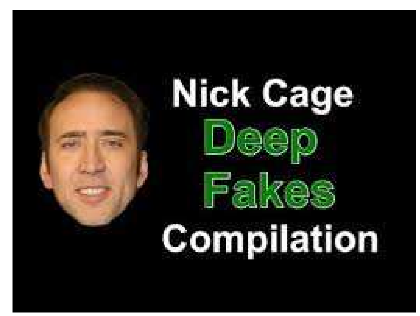


detect region splicing

FROM SEEING TO UNDERSTANDING

Can We Trust What We See?

2017: DeepFake https://en.wikipedia.org/wiki/Deepfake



https://gizmodo.com/insanely-accurate-lip-synching-tech-could-turn-fake-new-1796843610

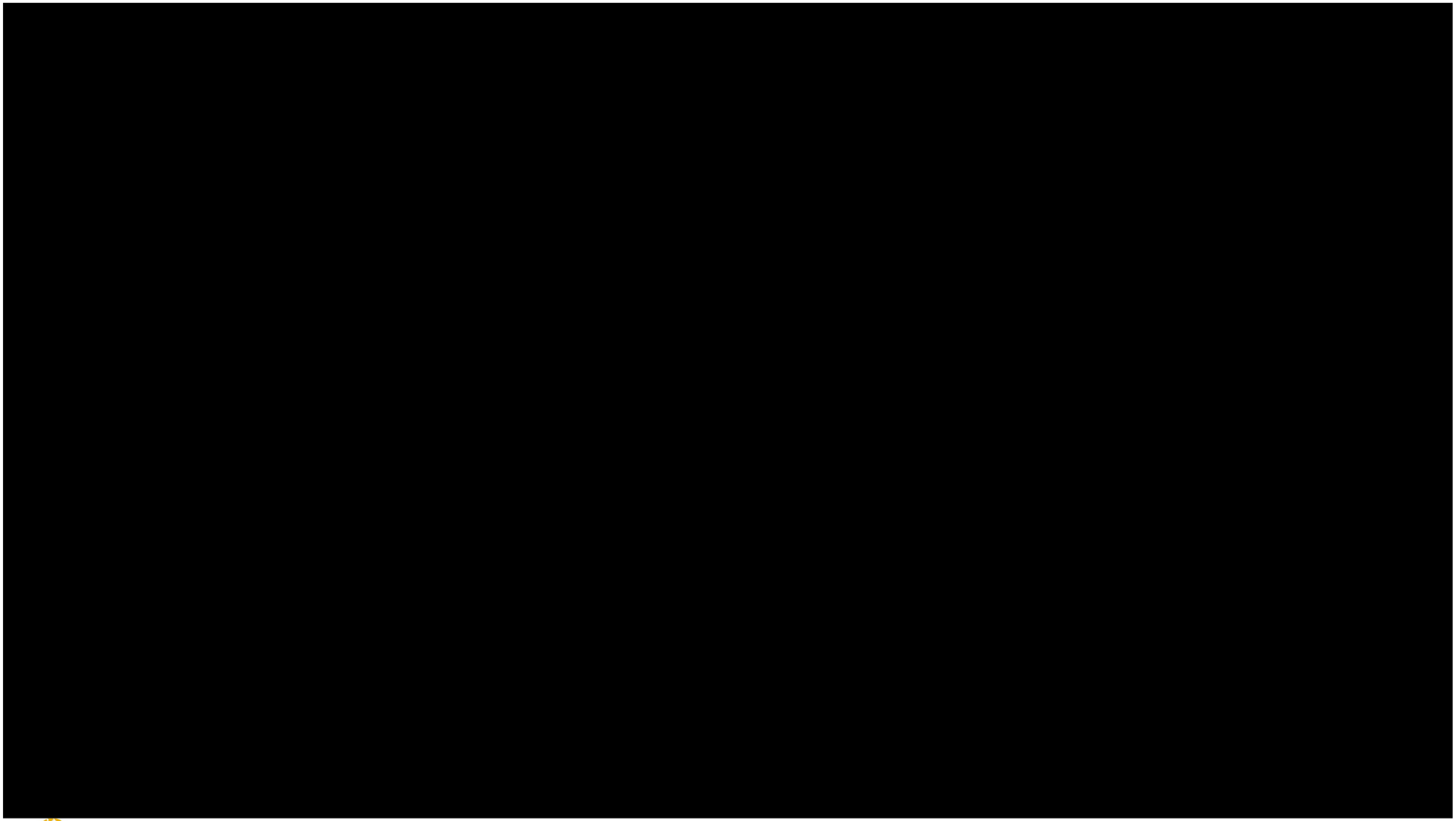
https://www.youtube.com/watch?v=BU9YAHigNx8

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Nick Cage DeepFake Videos

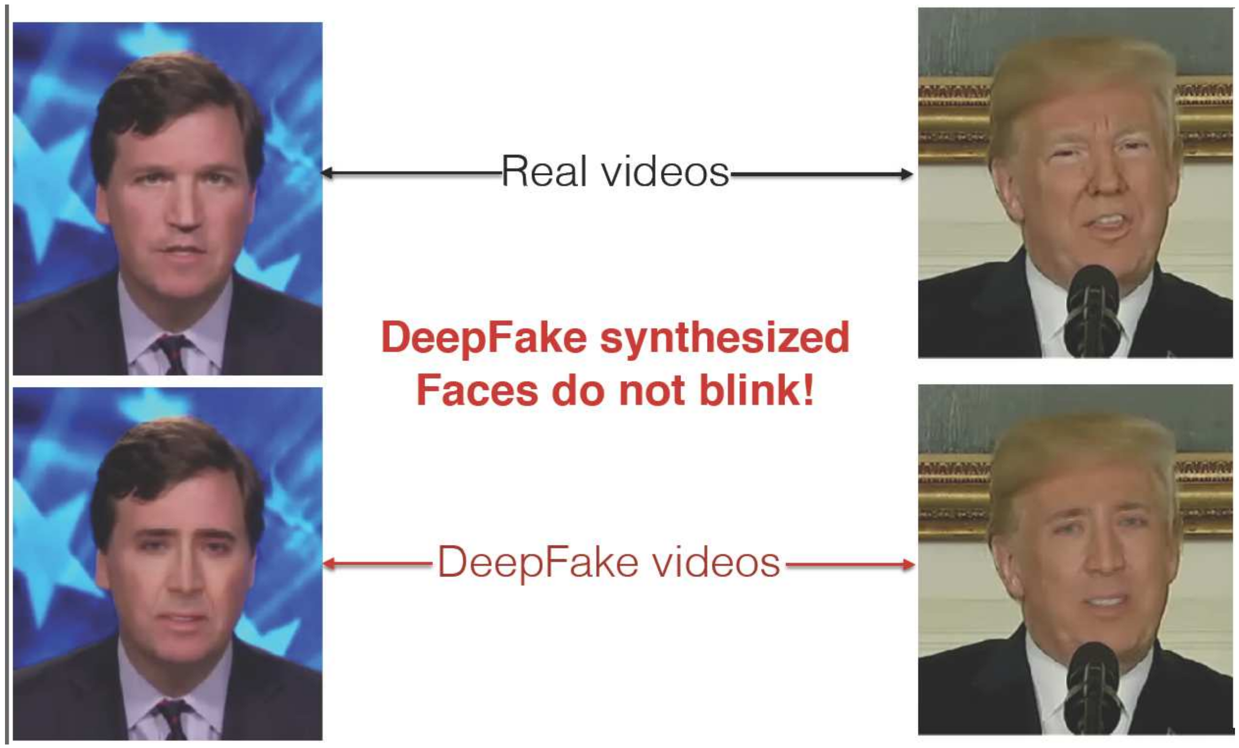
https://www.youtube.com/watch?v=BU9YAHigNx8

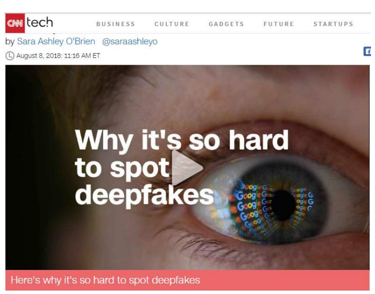


Fake Obama Created Using AI Video Tool - BBC News

https://www.youtube.com/watch?v=AmUC4m6w1wo

Detect DeepFake Videos



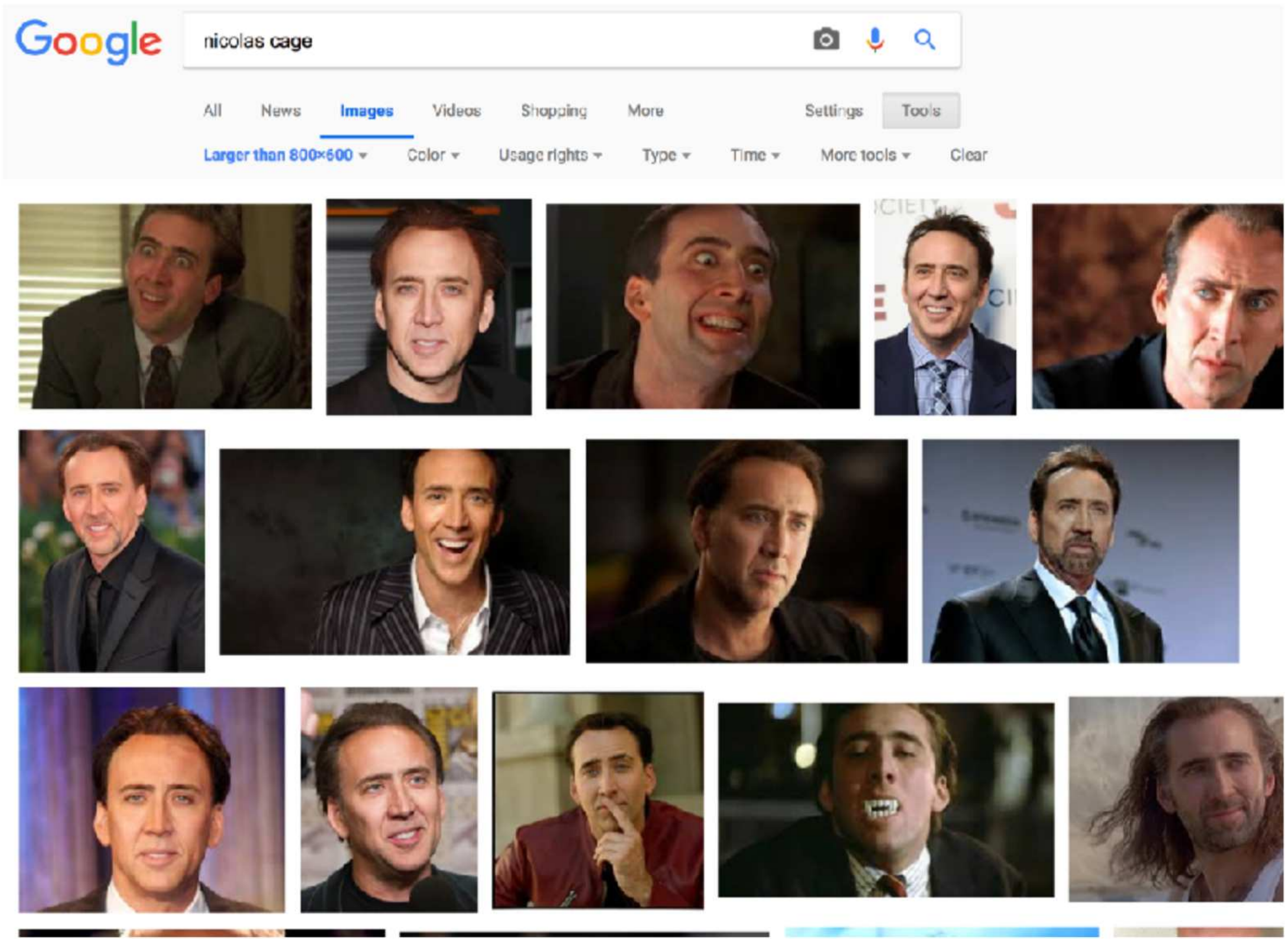


This work was recently featured in *CNN*, *The Guardian*, *Wired*, *The Register (UK)*, McClatchy, BBC News and several prominent publications.

[Yuezun Li, Ming-Ching Chang, Siwei Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking", WIFS 2018]

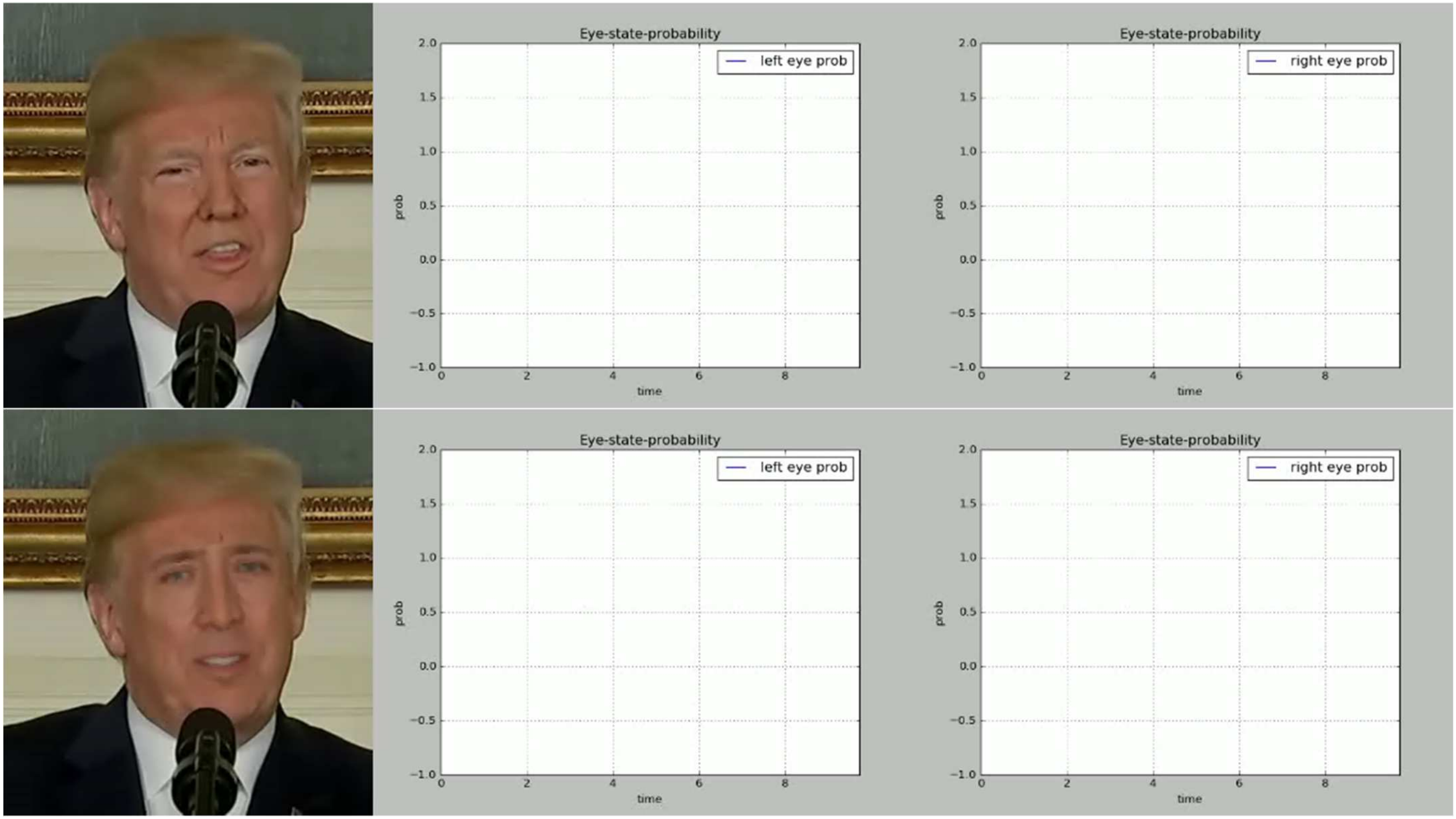
Detect DeepFake Videos

Reason: training data may not include closed eyes



[Yuezun Li, Ming-Ching Chang, Siwei Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking", WIFS 2018]

Detect DeepFake Videos



in ictu oculi (in the blink of the eye)

[Yuezun Li, Ming-Ching Chang, Siwei Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking", WIFS 2018]

UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

DARPA MediFor Program

https://www.darpa.mil/program/media-forensics

DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

EXPLORE BY TAG

ABOUT US / OUR RESEARCH / NEWS / EVENTS / WORK WITH US /

Defense Advanced Research Projects Agency > Program Information

Media Forensics (MediFor)

Dr. Matt Turek

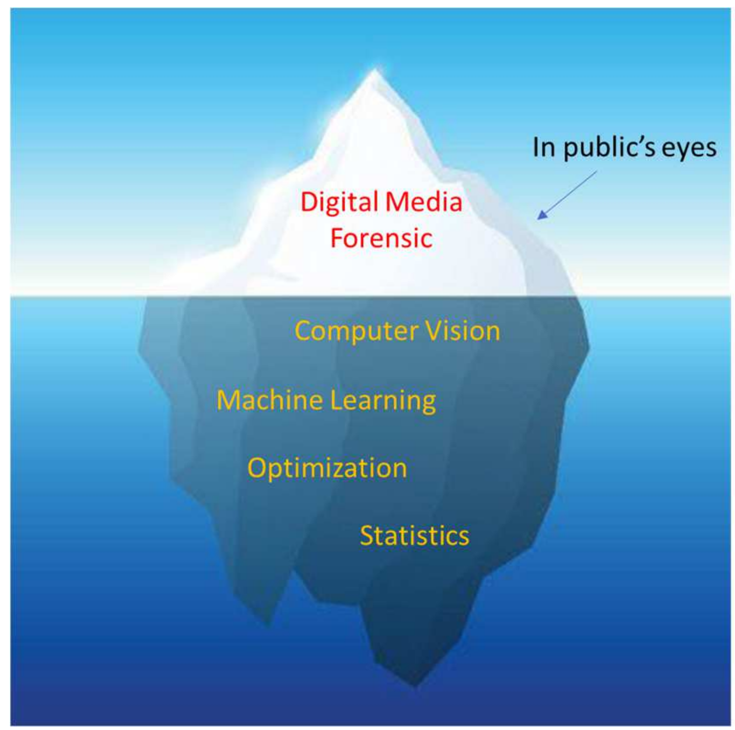
Historically, the U.S. Government deployed and operated a variety of collection systems that provided imagery with assured integrity. In recent years however, consumer imaging technology (digital cameras, mobile phones, etc.) has become ubiquitous, allowing people the world over to take and share images and video instantaneously. Mirroring this rise in digital imagery is the associated ability for even relatively unskilled users to manipulate and distort the message of the visual media. While many manipulations are benign, performed for fun or for artistic value, others are for adversarial purposes, such as propaganda or misinformation campaigns.

This manipulation of visual media is enabled by the wide scale availability of sophisticated image and video editing applications as well as automated manipulation algorithms that permit editing in ways that are very difficult to detect either visually or with current image analysis and visual media forensics tools. The forensic tools used today lack robustness and scalability, and address only some aspects of media authentication; an end-to-end platform to perform a complete and automated forensic analysis does not exist.

DARPA's MediFor program brings together world-class researchers to attempt to level the digital imagery playing field, which currently favors the manipulator, by developing technologies for the automated assessment of the integrity of an image or video and integrating these in an end-to-end media forensics platform. If successful, the MediFor platform will automatically detect manipulations, provide detailed information about how these manipulations were performed, and reason about the overall integrity of visual media to facilitate decisions regarding the use of any questionable image or video.

Digital Media Forensic

- Technologies for creating fake media advance rapidly and can cause serious impacts to society.
  - Latest Deepfake tech will make a whole-body dancing video by transfer the motion of one person to another in a video.
- Digital media forensics are catching up to control the negative effects of fake media.
- The rivalry between forgeries and forensics will continue for coming years.



UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Ethics of AI, Trustworthiness

Study of:

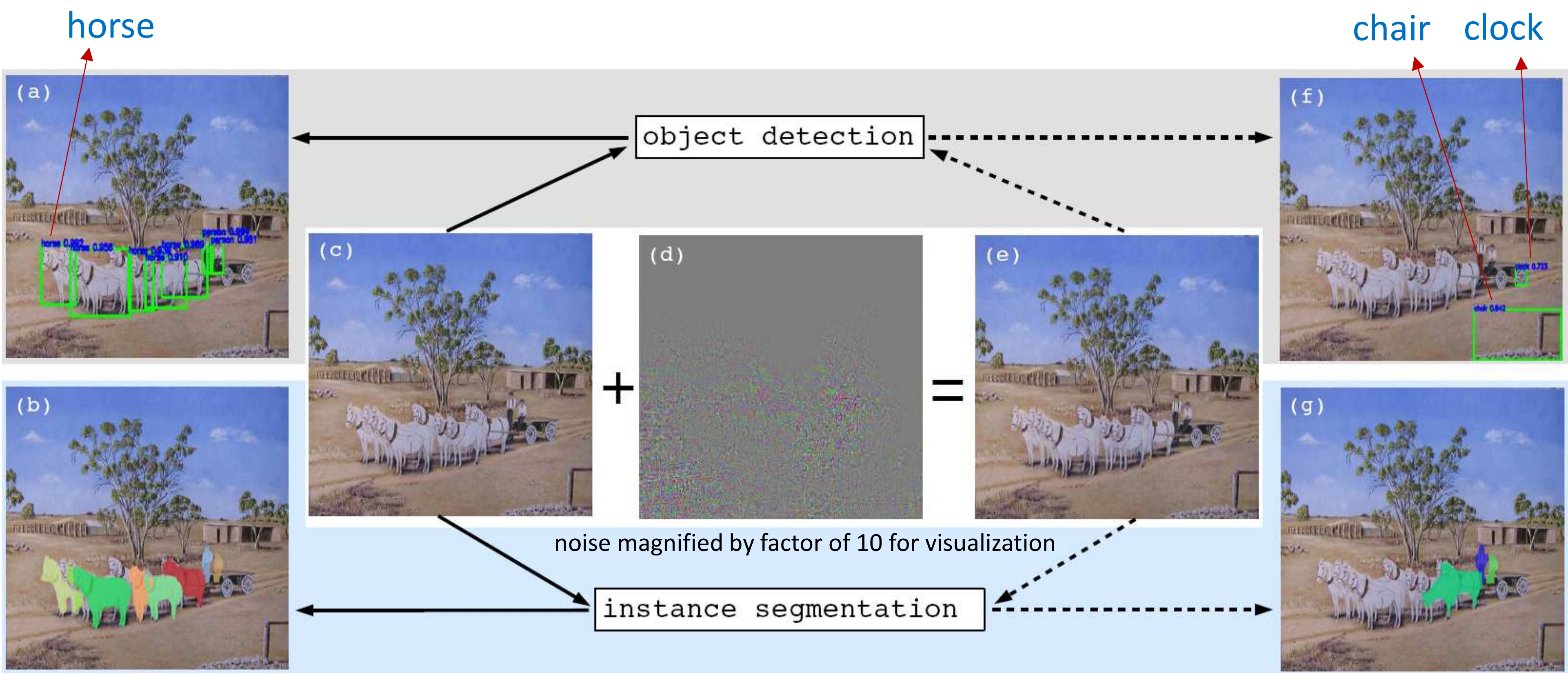
- Ethics of AI technologies
- Ways to improve the trustworthy of AI technologies

will be the next hot research topic.

Image from http://aiethicslab.com/

DNN Vulnerable to Adversarial Noises

Deep networks can be vulnerable to attacks from intentionally crafted perturbation imperceptible to human eyes.



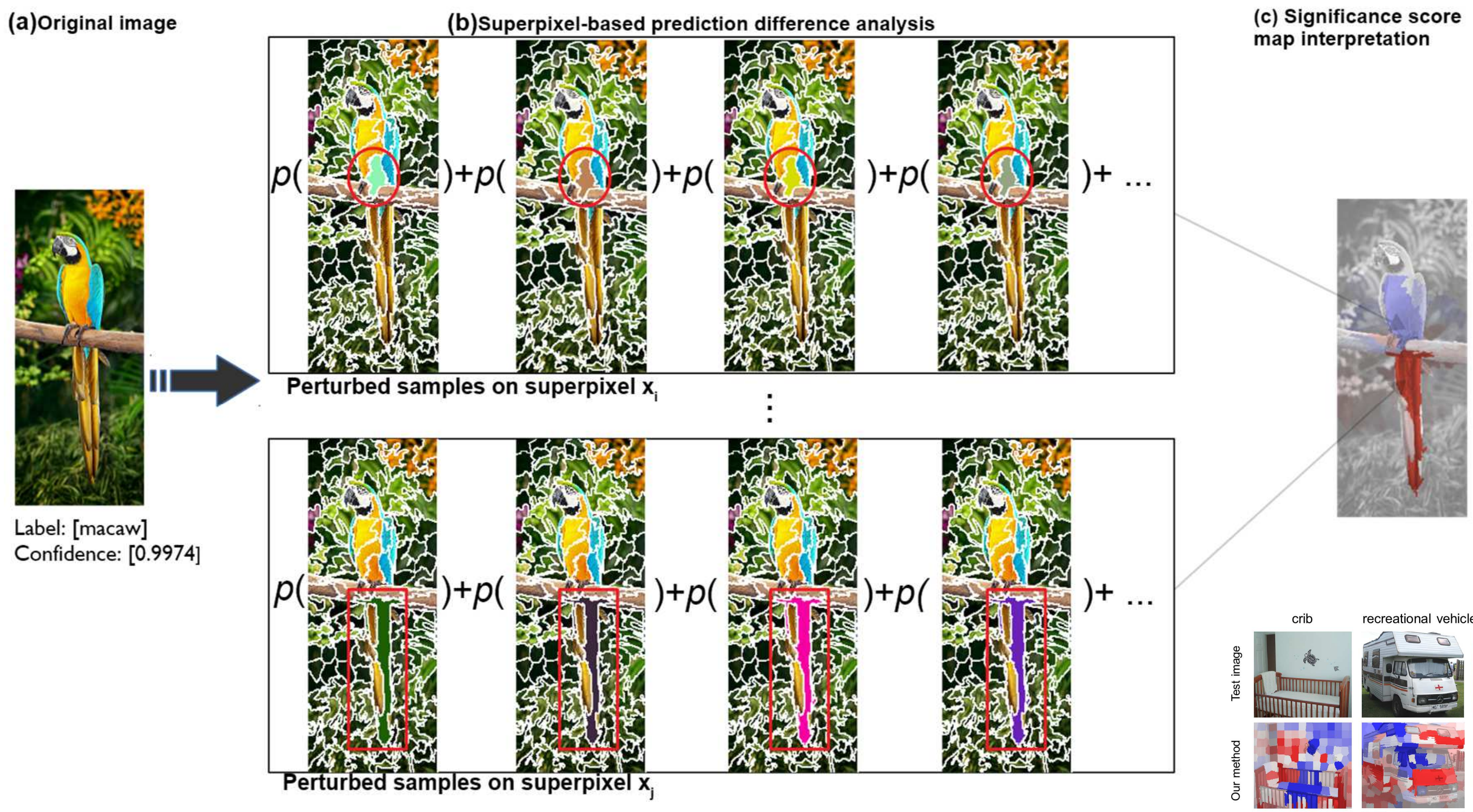
[Yuezun Li, Daniel Tian, Ming-Ching Chang, Xiao Bian, Siwei Lyu, "Robust Adversarial Perturbation on Deep Proposal-based Models", BMVC 2018]

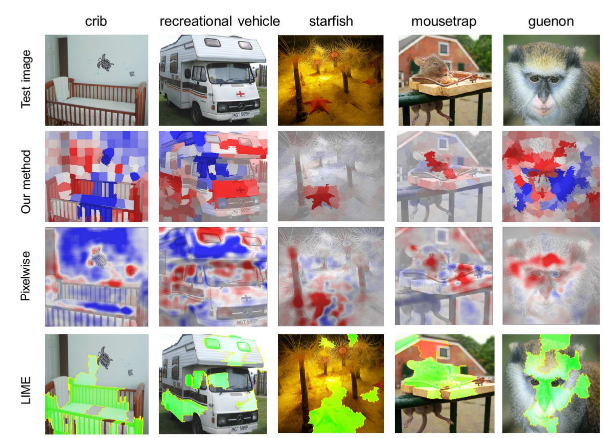
UNIVERSITY AT ALBANY
State University of New York

FROM SEEING TO UNDERSTANDING

Interpretable DNN

Black-box interpretation of how each image component contribute to the classification decision. Visualization shows if the DNN is trustworthy and possible ways to improve it.





[Yi Wei, Ming-Ching Chang, Yiming Ying, Ser Nam Lim, Siwei Lyu, "Explain Black-Box Image Classification Using Superpixel-Based Interpretation", ICPR 2018]

Conclusion

- AI Technologies will continue to expand, methods will continue to improve.
- Impact to life will continue to grow, but less-perceived.
- AI in aiding decision support will eventually dominate the decision making.
  - Let the number (data analytics) talk.
- Ethics and policy of AI will become important.
  - Technology will not be the only driving force, but culture, economy, politics, ...
- AI human-nature harmony
  - AI rationality vs. common sense learning
  - Theory of mind, consciousness for AI, computational linguistics

Questions?

Talk Summary

- A Smart Future Full of Technologies
- Computer Vision in Smart Transportation
  - DETRAC Dataset & Benchmark
  - AI City Challenges
  - AVSS T4S Workshops
- Visual Scenario Understanding
  - Detection, Tracking, Mapping, Analysis
  - Human Pose, Face, Group, Social Interaction, ...
- Industrial Video Analytics
  - Inspection, healthcare, Situational Awareness, Site Management, Transportation
- Digital Media Forensics
  - Ethics and trustworthiness of AI

Thank You