Virtual Vision:

Computer Vision in Virtual Reality

Demetri Terzopoulos

University of California, Los Angeles

Visual Modeling & Computing

Computer Graphics

- Synthesis:
  - *From mathematical models to images*
  - *Forward problem*

and Computer Vision

- Analysis:
  - *From images to mathematical models*
  - *Inverse problem*

Visual Surveillance

Visual surveillance is becoming ubiquitous

  - *London has roughly 4,000,000 cameras*

- Anti-Terrorism, deterring crime, etc.
- Effective visual coverage of large spaces require multi-camera systems
- Operator monitoring infeasible in large networks
- Need *autonomous networks of smart cameras*
- Smart cameras: Visual sensor nodes
  - *On-board processing, communication*

- Problem: **Large-scale** visual sensor network research is infeasible for most computer vision and sensor networks researchers

The difficulty of Doing Large-Scale Visual Sensor Networks Research

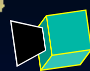
Deploying large-scale camera networks in extensive public spaces for research purposes:

- Very costly
- Privacy and legal issues
- Hardware-related technical challenges

Infeasible for most computer vision researchers

Virtual Vision

Visually and behaviorally realistic simulators for designing and evaluating machine vision systems



Virtual Camera

Camera Model:
pan, tilt, zoom, camera jitter, color response, lens distortions, etc.

High-Level Control
camera control, assignment, handover, etc.

Environment Models:
geometry, texture, illumination

Pedestrian Models:
appearance, movement, behavior

Reality Emulator
Virtual Penn Station

Machine Vision
tracking, recognition, etc.

Virtual Vision: A Tool for Visual Sensor Network Research

High-Level Camera Control

Virtual Vision		Real Vision
Visual Sensing	⇔	Visual Sensing
Synthetic Video	⇔	Real Video
Virtual Camera Network	⇔	Physical Camera Network
Synthetic World (Reality Emulator)	⇔	Real World

Virtual Vision **Real Vision**

2007 PhD Thesis of Faisal Qureshi University of Toronto

Publications:

- Chapters in "*Distributed Video Sensor Networks*", Bhanu, et al., Springer, 2011
- 2008 Proceedings of the IEEE
- 2006 ACM Multimedia Systems Journal
- 2009 Third IEEE/ACM Intl. Conf. on Distributed Smart Cameras *(ICDSC)*
- 2008 11th Communications and Networking Simulation Symposium *(CNS)*
- 2008 ACM Symposium on Virtual Reality Software and Technology *(VRST)*
- 2007 First IEEE/ACM Intl. Conf. on Distributed Smart Cameras *(ICDSC)*
- 2007 IEEE/ACM Intl. Conf. on Dist. Comp. in Sensor Systems *(DCOSS)*
- 2007 IEEE Conf. on Computer Vision and Pattern Recognition *(CVPR)*
- 2006 ACM Intl. Wksp. on Distributed Smart Cameras *(DSC)*
- 2005 ACM Wksp. on Video Surveillance and Sensor Networks *(VSSN)*
- 2005 IEEE Intl. Wksp. on Visual Surveillance *(VS-PETS)*

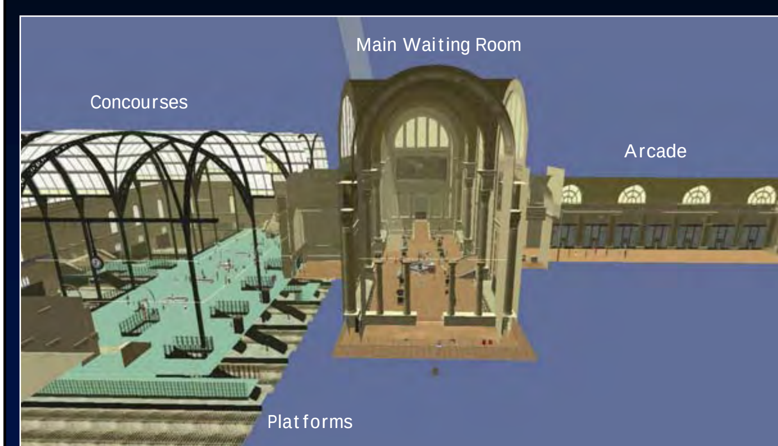
2006 PhD Thesis of Wei Shao New York University

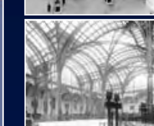
Publications:

- 2007 Graphical Models
- 2006 SAE Transactions Journal
- 2006 Int. Conf. on Intelligent Virtual Agents (IVA)
- 2005 SIGGRAPH/EG Symposium On Computer Animation (SCA)

Autonomous Pedestrian Simulation

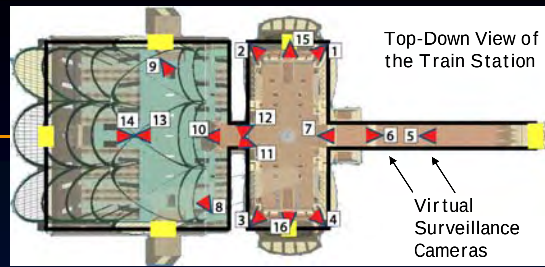
Environmental Model of Original Penn Station in NYC

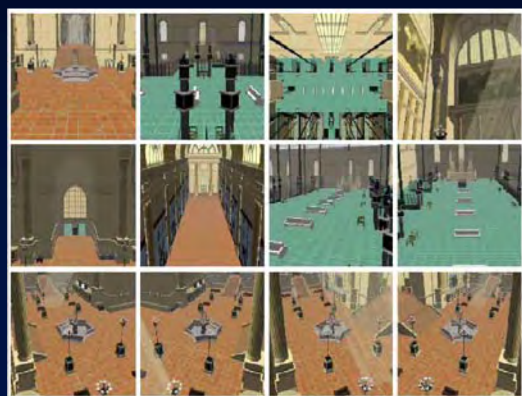




Virtual Vision

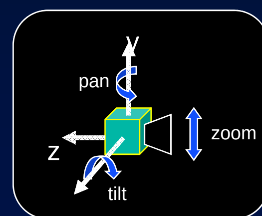
Visual sensor networks





Virtual Surveillance Camera Video Feeds

Active Pan-Tilt-Zoom Camera

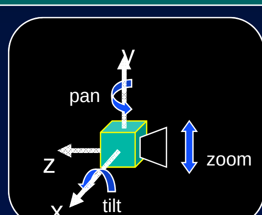


Anatomy of a Smart Camera Node

Communication subsystem: message passing to neighboring nodes

Task "relevance" computation framework

Vision routines: pedestrian tracking



active pan-tilt-zoom camera

Decision logic

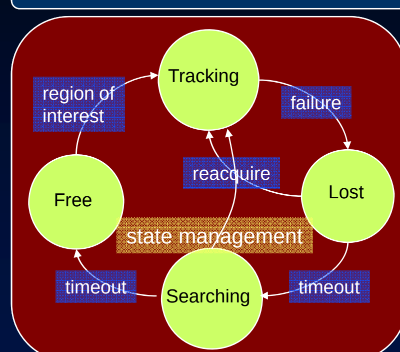


Image driven reactive behaviors: fixation and zooming PD controllers

Synthetic Video

Model characteristics of real CCTV video

- Camera color response
- Camera distortion
- Camera noise – detector noise, data drop-out noise
- Compression artifacts
- Interlacing artifacts

Camera Noise

Compression Artifacts

Interlacing Artifacts

Computer Vision Emulation Using Synthetic Video

Pedestrian detection

- Background subtraction using a learnt background model

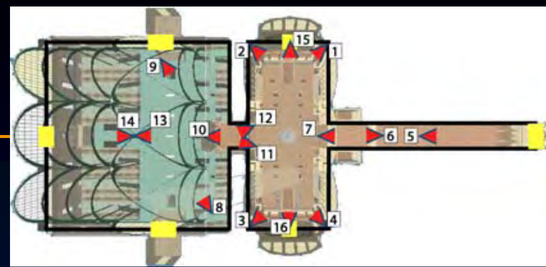
Computer Vision Emulation Using Synthetic Video

Appearance based pedestrian tracker

- Swain & Ballard 91

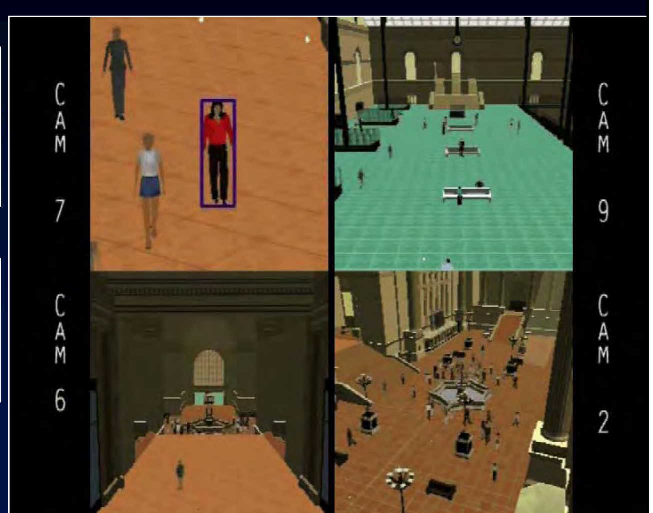
Visual Tracking of Pedestrians

Low-level computer vision



Automatic, Persistent Multicamera Surveillance:

Active PTZ Camera Assignment and Grouping



Network Model

- Does not require camera or network calibration
  - *Can take advantage of calibration information, if available*
  - *Ad hoc deployment*
- Camera grouping is a strictly local negotiation
  - *Typically camera groups are spatially local arrangements*
- Camera groups are dynamic arrangements
- Camera handoffs occur naturally during negotiations
- Camera nodes can be added/removed during a task

Network Model

- It can gracefully handle node and message failures
- Even assuming perfect sensing, the proposed model can still fail if
  - *a significant number of messages are lost*
  - *catastrophic node failure*
  - *group evolution can't keep up with a fast-changing observation task*
- Scalability
  - *Small group sizes*
  - *Conflict resolution is viable as long as the number of relevant sensors for each task remains low (< 10)*
- Optimal sensor assignment

Benefits of Virtual Vision for Camera Networks Research

- Emulates the characteristics of a physical vision system
- Flexibility during system design and evolution
- Readily available ground truth
- Online operation and testing
- No legal impediments
- No special hardware
- Repeatability
- Inexpensive

QuickTime™ and a
decompressor
are needed to see this picture.

Camera Sensor Network #1

Active PTZ Camera Scheduling

Scheduling Active PTZ Cameras

Scheduling problem

- Given n PTZ cameras, and m pedestrians, persistently observe every pedestrian using one PTZ camera at a time

Goal

- Observe as many pedestrians for as long as possible

Virtual Active Camera Scheduling

Passive wide-FOV cameras

- Calibrated
- Pedestrian localization through triangulation

Active PTZ cameras

- Un-calibrated
- Learn a coarse mapping between 3D locations and internal pan-tilt settings

Reliable pedestrian identification in different cameras via appearance based signatures

Calibrated passive cameras at the four corners of the waiting room in the virtual train station

Scheduling Active PTZ Cameras

Number of Pedestrians > Number of active cameras

Task active cameras to observe pedestrians in the scene

Active PTZ camera

Passive wide-FOV camera

Active Camera Scheduling Strategy

Camera assignment via weighted round robin

First come, first served tie breaking

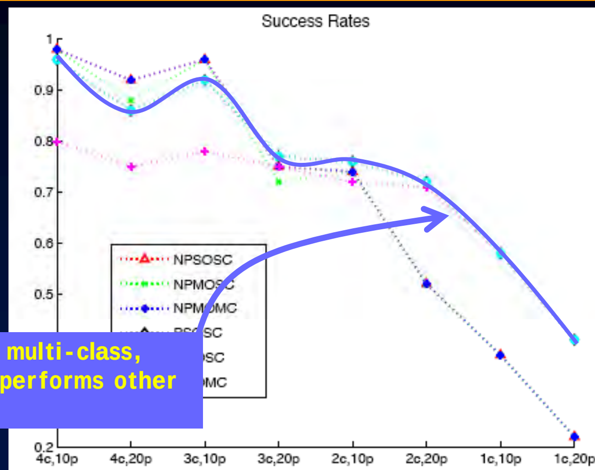
Multiple observation

Preemption

Close-up snapshots captured
by active PTZ cameras

Active Camera Scheduling Results

- *Preemption (P)*
- *No preemption (NP)*
- *Single observation (SO)*
- *Multiple observations (MO)*
- *Single-class model (SC)*
- *Multi-class model (MC)*



Multiple observations, multi-class, preemption scheduler outperforms other variants

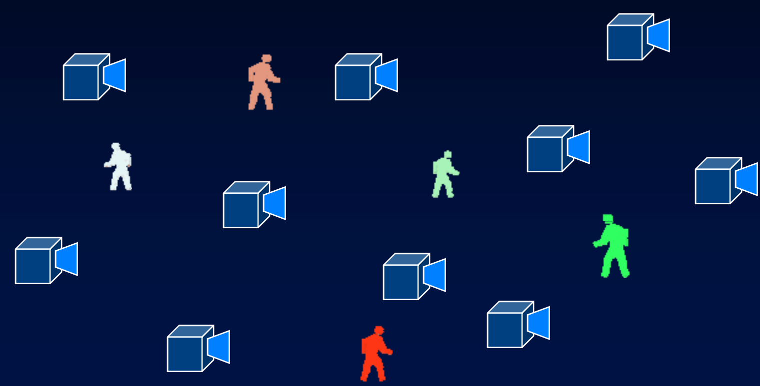
Up to 4 Cameras; 10, 20 Pedestrians

Camera Sensor Network #2

Active PTZ Camera Assignment and Grouping

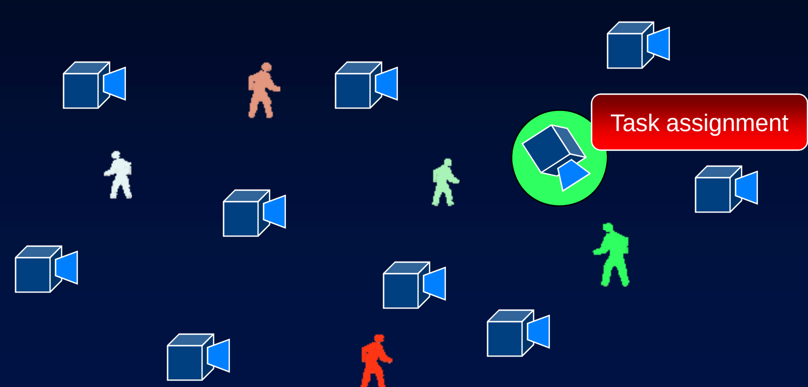
Vision and Goal

Ad hoc deployment



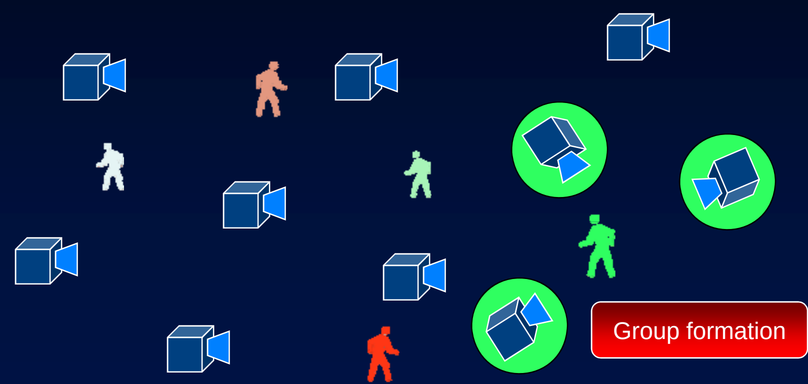
Objective

Cameras work towards common sensing goals



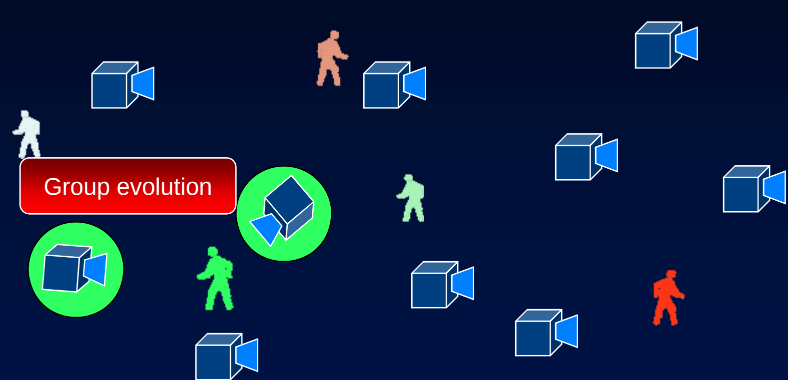
Vision and Goal

Cameras work towards common sensing goals



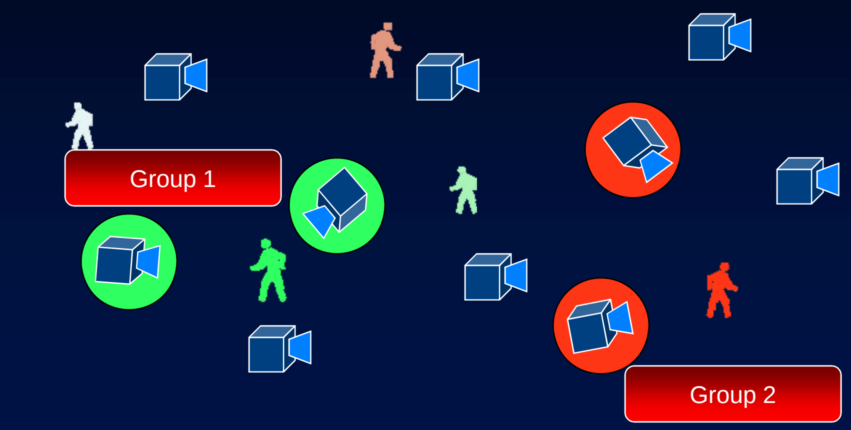
Vision and Goal

Cameras work towards common sensing goals



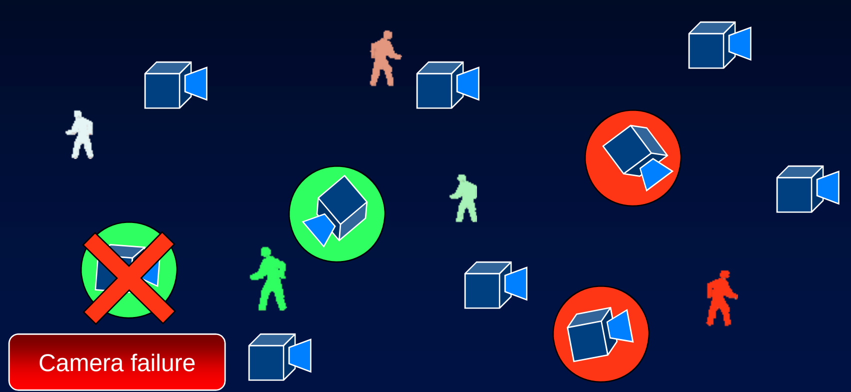
Vision and Goal

Cameras work towards multiple common sensing goals



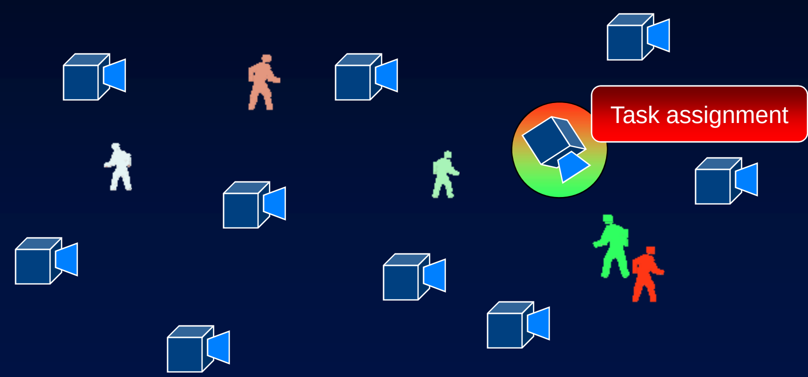
Vision and Goal

Cameras work towards multiple common sensing goals



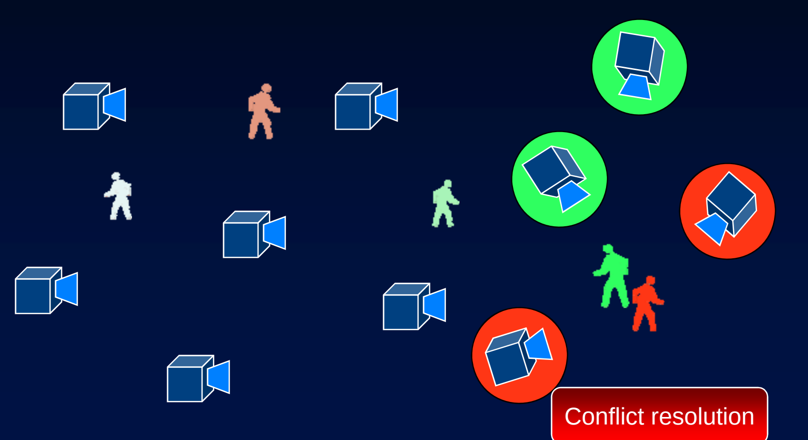
Vision and Goal

Cameras work towards multiple common sensing goals



Vision and Goal

Cameras work towards multiple common sensing goals



Camera Grouping and Reassignment

Active PTZ Camera Assignment and Grouping

Camera selection, grouping, and handoff via an auction model

- Announcement/Bidding/Selection
- ContractNet
  - *Smith, 1983*

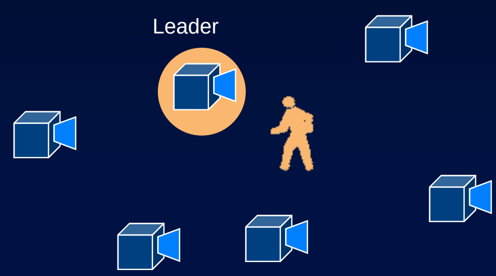
Conflict resolution within a Constraint Satisfaction Problem framework

- Partially distributed

A Camera can only perform a single task at any given time

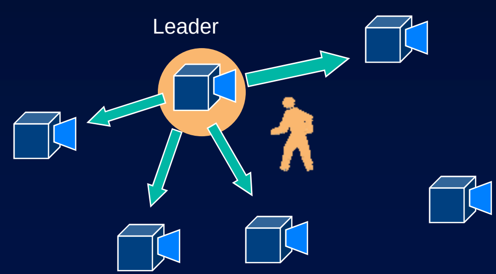
Camera Grouping: Announcement

- Start with a single camera that is tasked to observe a person



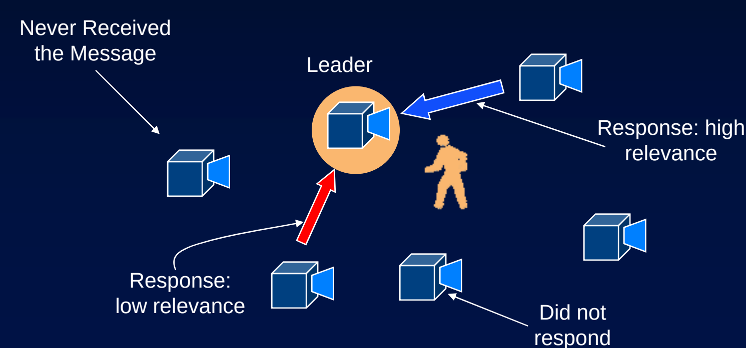
Camera Grouping: Announcement

- Seeks out other cameras in the vicinity to form a group to help it with the observation task



Camera Grouping: Bidding

- One or more cameras that receive the task announcement respond with their relevance values

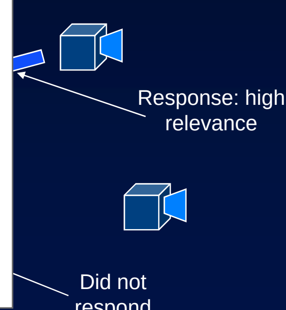


Camera Grouping: Bidding

- One or more cameras that receive the task announcement respond with their relevance values
- Relevance encodes how successful a camera will be at an observation task

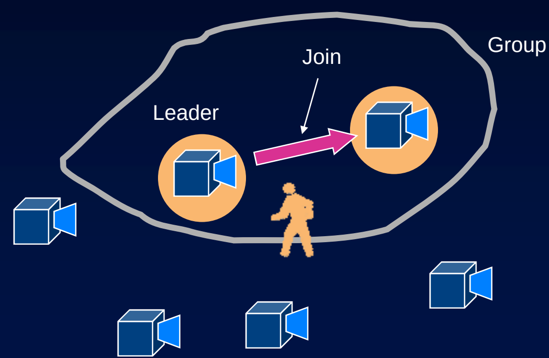
$$r = \begin{cases} \exp\left(-\frac{(c-1)^2}{2{\sigma_c}^2} - \frac{(\theta-\hat{\theta})^2}{2{\sigma_\theta}^2} - \frac{(\alpha-\hat{\alpha})^2}{2{\sigma_\alpha}^2} - \frac{(\beta-\hat{\beta})^2}{2{\sigma_\beta}^2}\right) & \text{when } s = \text{free} \\ \frac{t}{t+\gamma} \quad \text{when } s = \text{busy} & \end{cases}$$

$$\begin{aligned} \textit{Status} &= s \in \{\text{busy}, \text{free}\} \\ \textit{Quality} &= c \in [0, 1] \\ \textit{Fov} &= \theta \in [\theta_{\text{min}}, \theta_{\text{max}}] \text{ degrees} \\ \textit{XTurn} &= \alpha \in [\alpha_{\text{min}}, \alpha_{\text{max}}] \text{ degrees} \\ \textit{YTurn} &= \beta \in [\beta_{\text{min}}, \beta_{\text{max}}] \text{ degrees} \\ \textit{Time} &= t \in [0, \infty) \text{ seconds} \end{aligned}$$



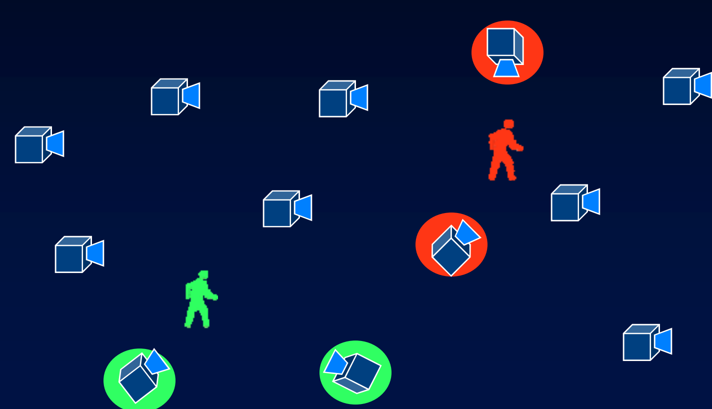
Camera Grouping: Selection

- After the leader gets relevance messages from neighboring cameras, it selects suitable cameras to join the group



Conflict Detection

- A conflict is detected when multiple tasks require the same camera node to proceed successfully

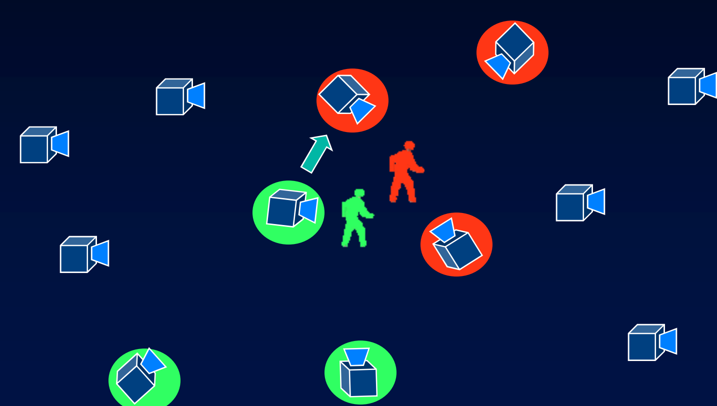


Conflict Detection

A Red group member receives a recruit query from Green group

Conflict Detection

A Red group member receives a recruit query from Green group



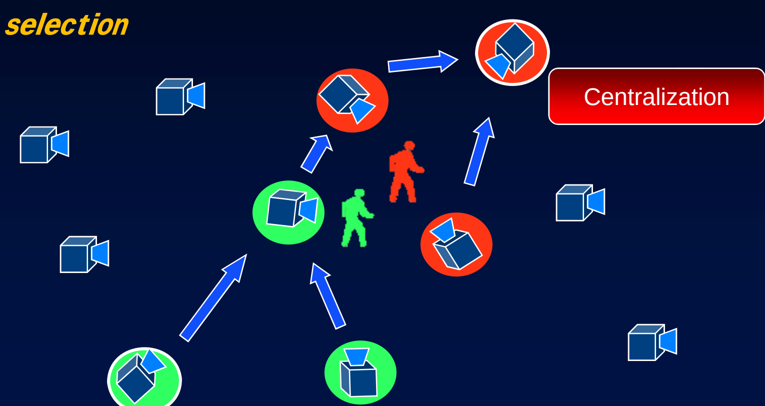
Conflict Detection

A Red group member receives a recruit query from Green group

Conflict Detection

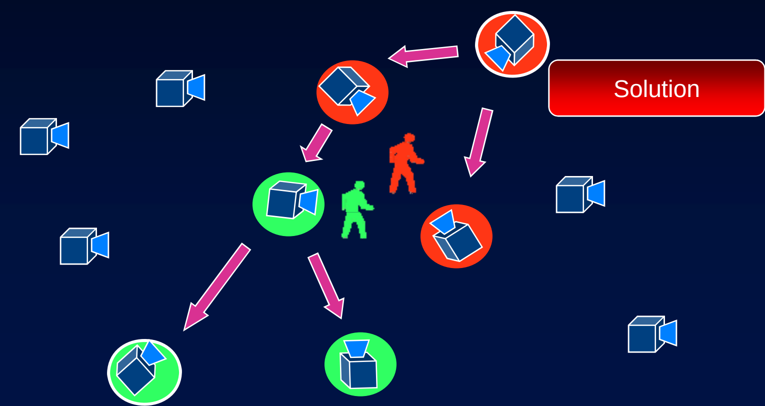
Nodes belonging to both groups send information to one of the leaders

Leader selection



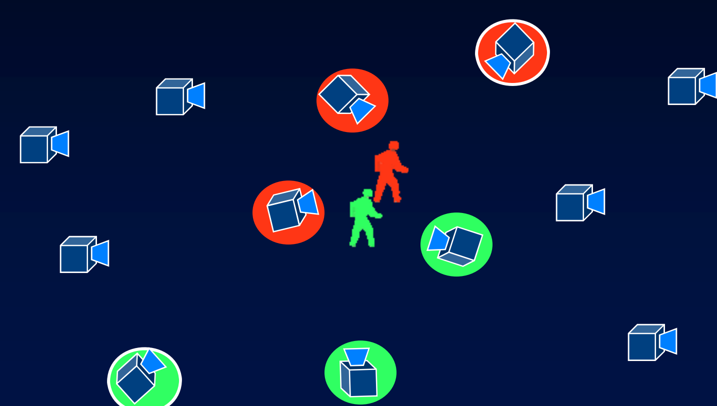
Conflict Detection

The resulting node (camera) assignment is sent to the individual nodes



Conflict Detection

The resulting node (camera) assignment is sent to the individual nodes



Camera Sensor Network #3

Planning for PTZ Camera Control

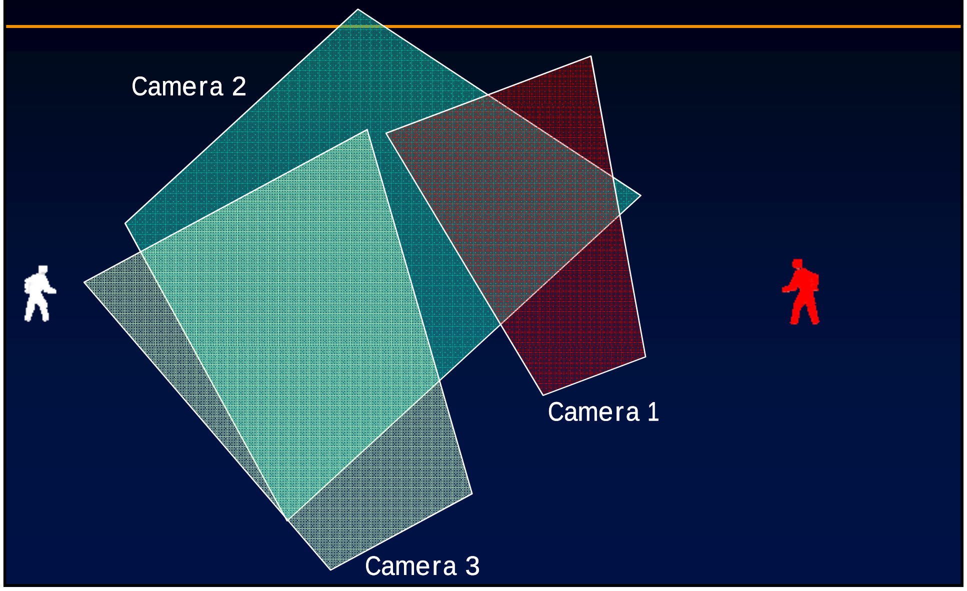
Planning for PTZ Camera Control

We formulate PTZ camera control as a planning problem whose solution achieves optimal *camera utilization w.r.t. to a predefined observational goal*

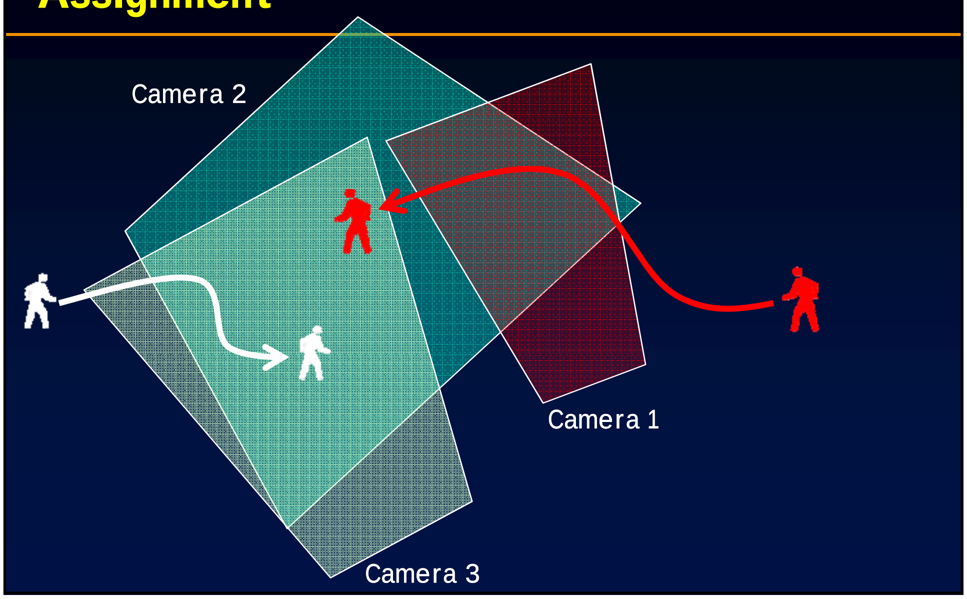
⬇

Achieve seamless closeup video of multiple pedestrians during their presence in a designated area

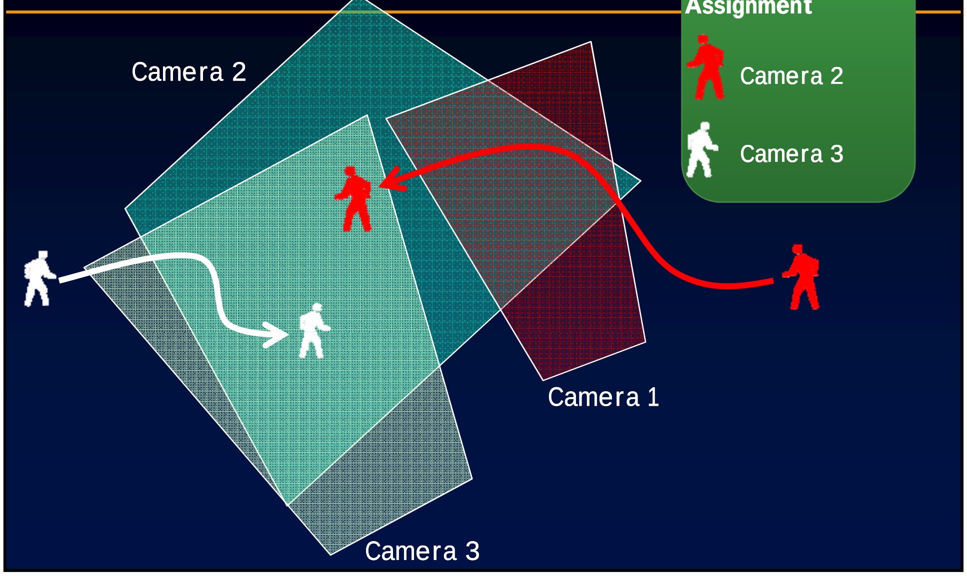
Proactive PTZ Camera Assignment



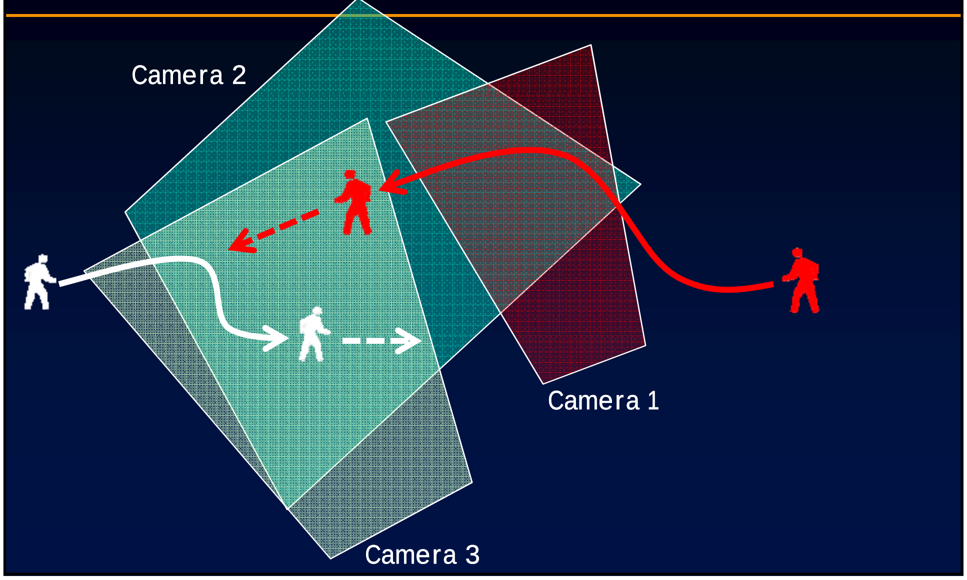
Proactive PTZ Camera Assignment

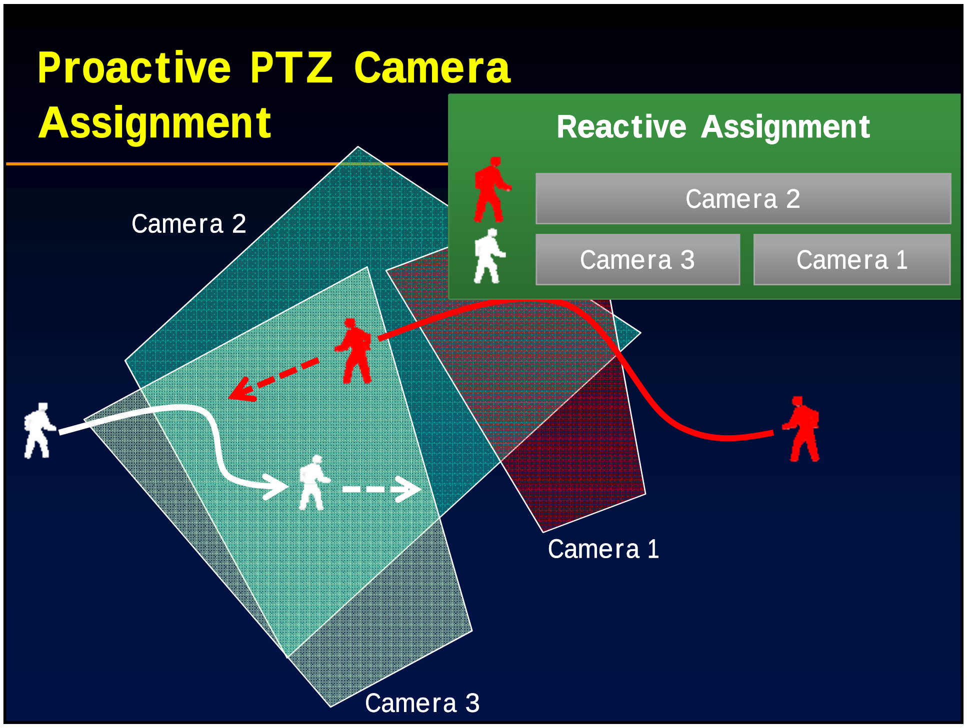


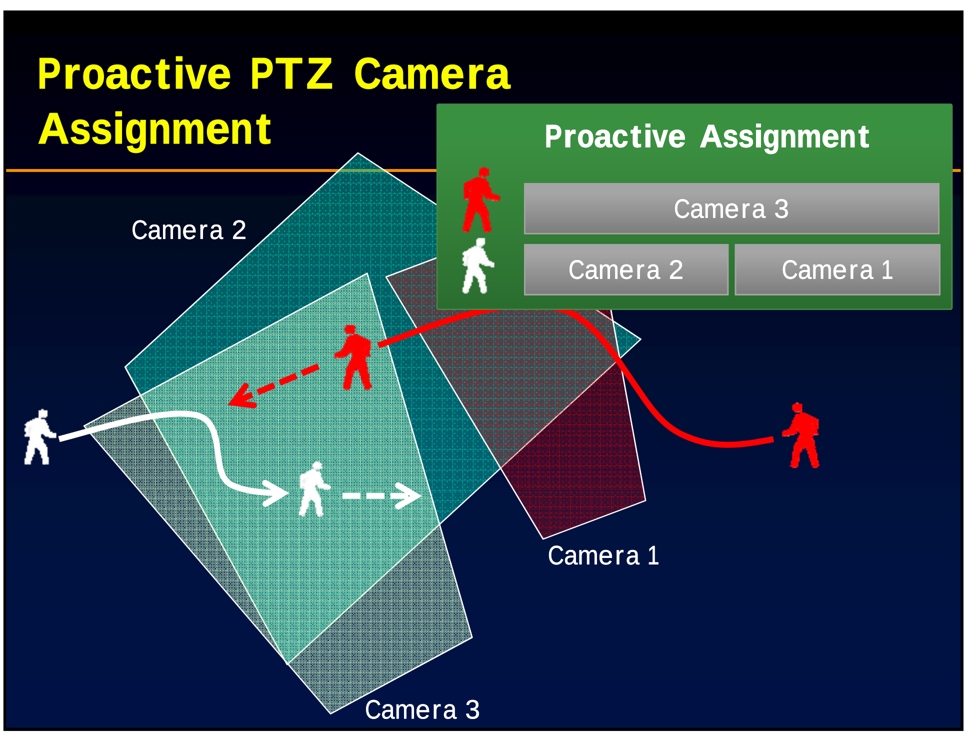
Proactive PTZ Camera Assignment



Proactive PTZ Camera Assignment




Proactive PTZ Camera Assignment
Reactive Assignment
Camera 2
Camera 3
Camera 1
Camera 2
Camera 1
Camera 3


Proactive PTZ Camera Assignment
Proactive Assignment
Camera 3
Camera 2
Camera 1
Camera 2
Camera 1
Camera 3

States and Actions

Planning problems are characterized by

- States
- Actions
- Goals

State space

Search for the best state sequence or action sequence

State at time *t*

***The status of every PTZ camera at time* t**

- Free
- Acquiring pedestrian
- Recording pedestrian

Action taken at time *t*

Actions available to a single camera

- Continue doing whatever it is doing before
- Stop recording and be Idle
- Start Acquiring a pedestrian
- Start Recording a pedestrian

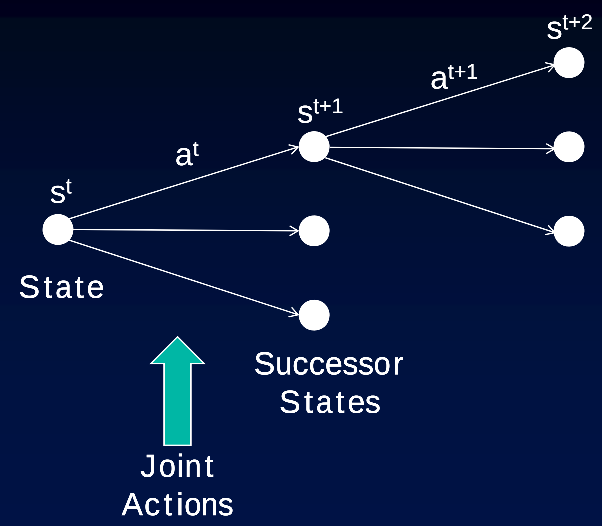
Joint Action taken at time *t*

Joint *Action taken by cameras at time* t

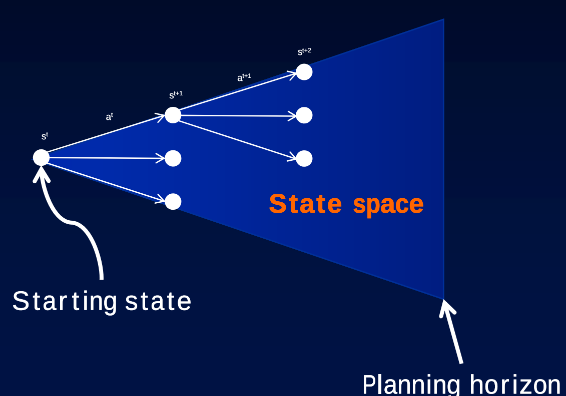
- Action taken by every PTZ camera at time *t*

Camera 1 is Acquiring pedestrian 1
Camera 2 is Continuing recording pedestrian 3
Camera 3 is Recording pedestrian 2

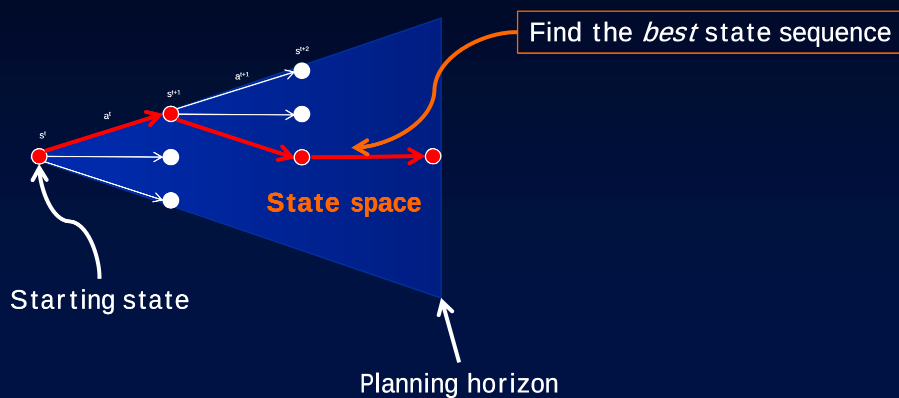
Planning Ahead for PTZ Camera Assignment



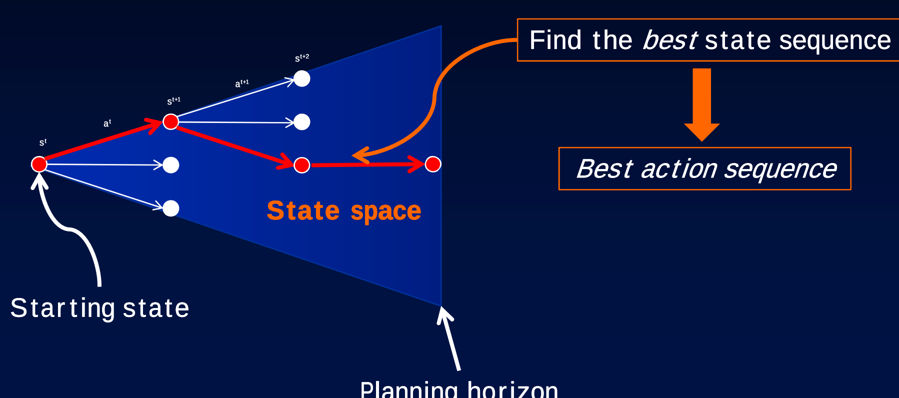
Planning Ahead for PTZ Camera Assignment



Planning Ahead for PTZ Camera Assignment



Planning Ahead for PTZ Camera Assignment



Finding the Best State Sequence

State sequence with the highest probability of success with respect to a given goal

↓

Each designated pedestrian must be viewed by at least one PTZ camera at all times

Finding the Best State Sequence

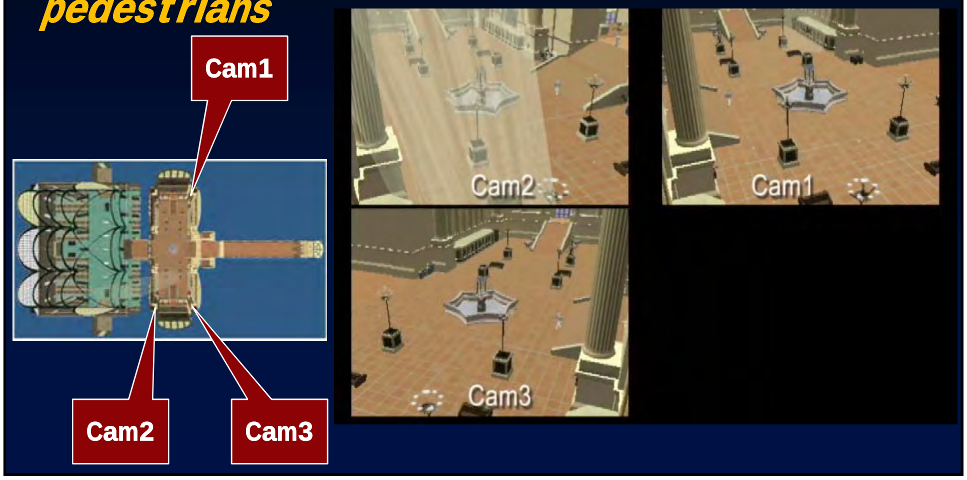
Success probability (Quality) of a state sequence

↳ **Success probability of individual states**

↳ **How successful individual PTZ cameras are at carrying out the tasks assigned to them**

Scenario 5

7 PTZ cameras are tasked to observe two pedestrians



Multi-Human Simulation

Autonomous Pedestrians

Self-Animating Virtual Humans in a Large-Scale Indoor Urban Space

Autonomous pedestrian simulation

Concourses

Main Waiting Room

Virtual Train Station

Autonomous Pedestrians

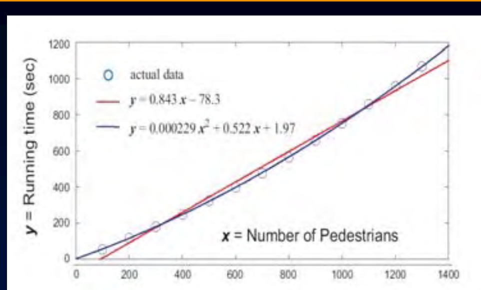
PhD thesis work of Wei Shao

Artificial life modeling approach

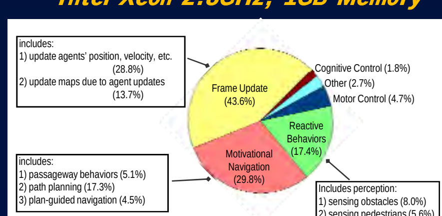
- Decentralized, autonomous, highly capable ***individuals***
- Comprehensive human model with motor, perceptual, behavioral, and cognitive components
- Hierarchical environment model

Each Pedestrian is a Capable Behavioral / Cognitive Individual

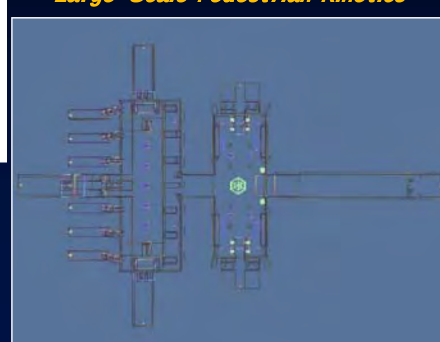
Performance



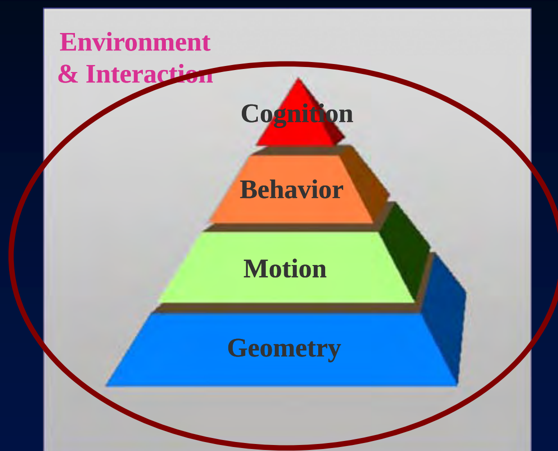
Intel Xeon 2.8GHz, 1GB Memory



Large-Scale Pedestrian Kinetics



Architecture of Pedestrian Model



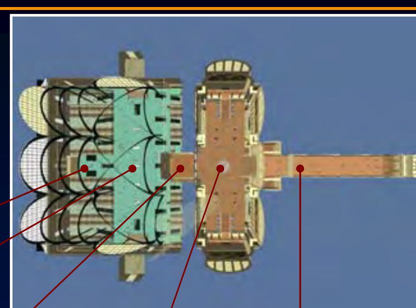
Penn Station Model

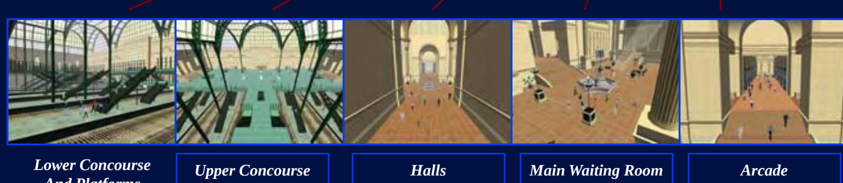
200m×150m×20m

43 regions

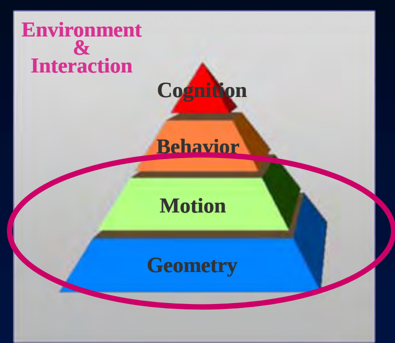
>500 objects (stairs, ticket booths, portals, …)

90MB memory





Geometry and Motor Control



DI-Guy (Boston Dynamics, Inc.)

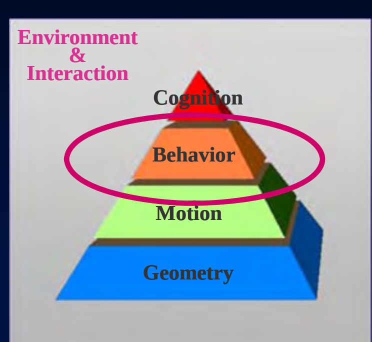
- Customized motion repertoire
- More responsive motion transitions

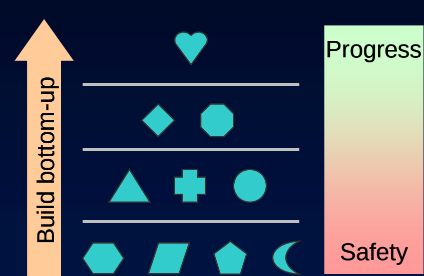
Motor control interface

- Verify motor control commands
- Hide underlying details

Behavior

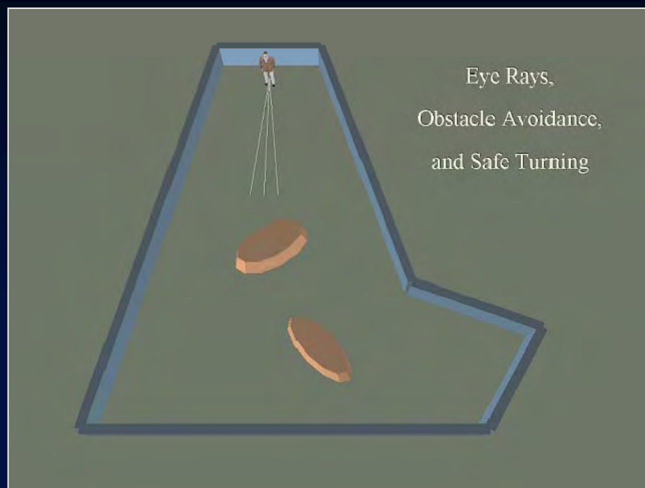
An ethological approach





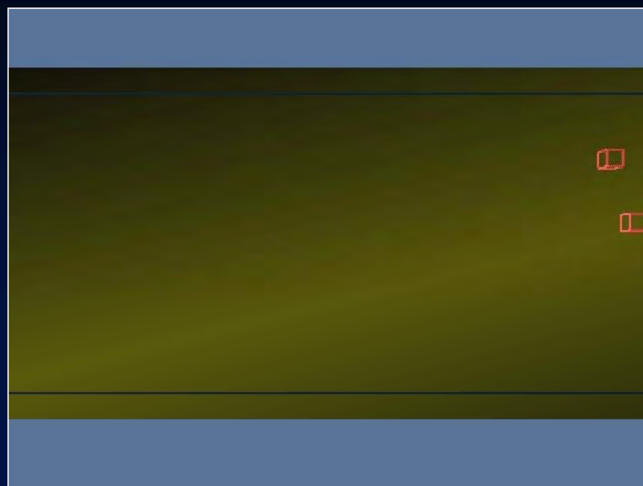
Behavior modules as building blocks

Perception-Guided Obstacle Avoidance



Perception-Guided Path Navigation

Corridor Navigation



Corridor Navigation

Other Motivational Behaviors

Select an unoccupied seat and sit down

Queue at ticket booths and purchase tickets

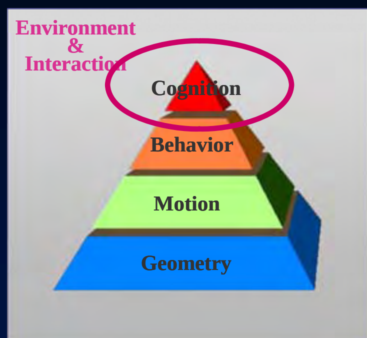
Approach artists and watch performances

Approach vending machines and purchase food

Social Conventions

Stay to the Right

Cognition



Global Path Planning

- Iteratively set intermediate goals at different levels
- Re-plan continuously

Heuristics

- Divide and conquer
- Think globally and act locally
- Be flexible

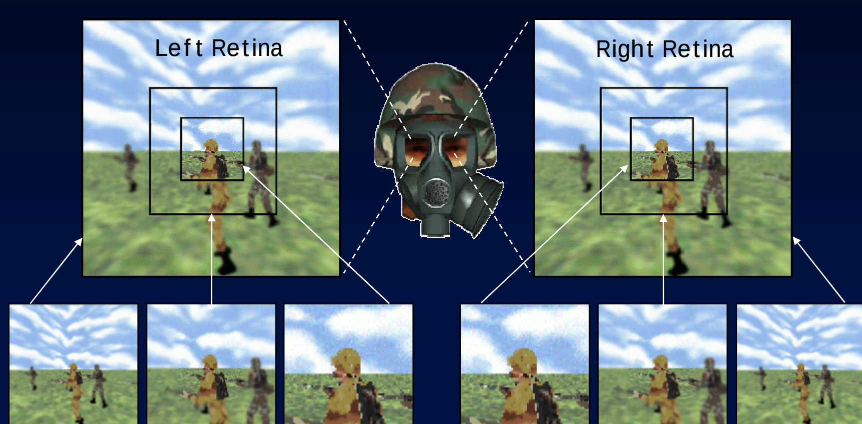
Related Virtual Vision Work

Active vision in (virtual) humans (and lower animals)

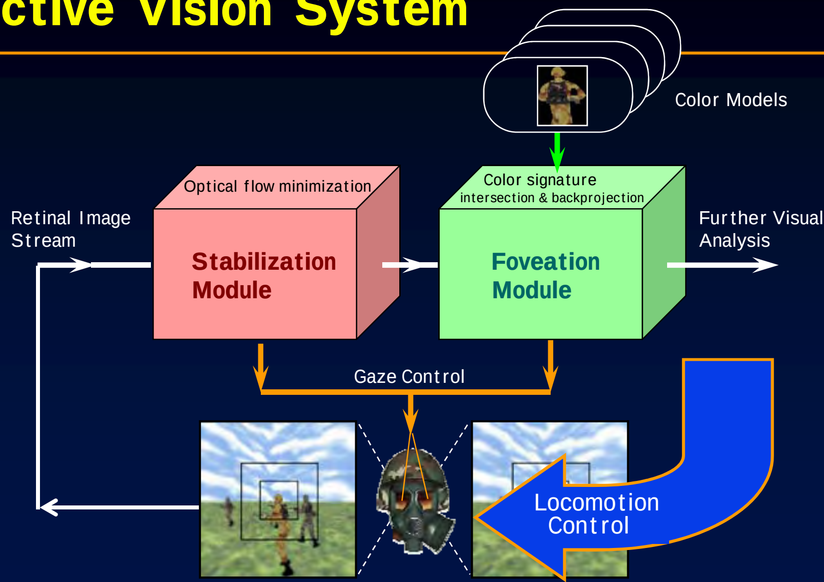
- 1999 PhD thesis of Tamer Rabie,
  University of Toronto

Visual Perception in Virtual Humans

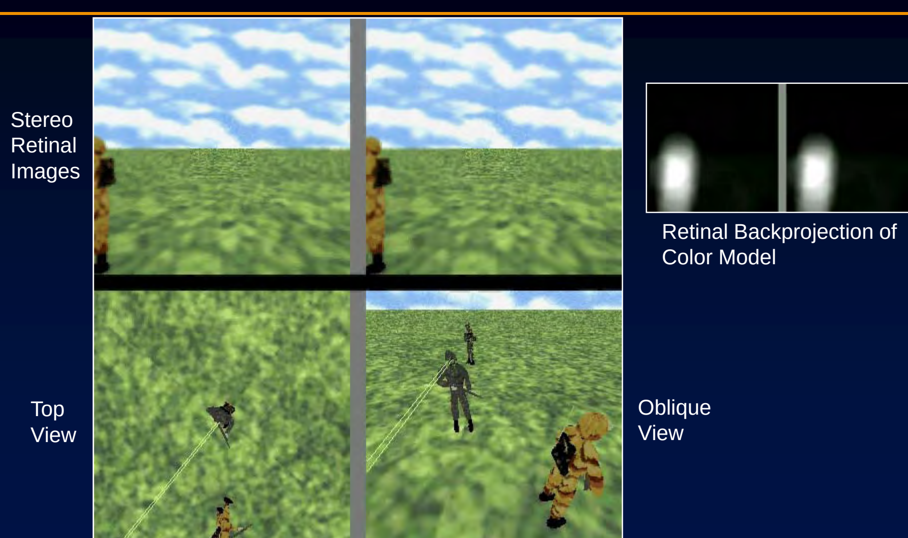
Virtual Eyes



Active Vision System



Visual Tracking and Sensorimotor Control



Conclusion

Advanced simulators are useful for computer vision research

- Particularly for research in intelligent, autonomous visual sensor networks
- Acceleration of the design/test cycle and scientific method

The gap between virtual reality and physical reality will continue to close over time

- The two should be indistinguishable in the long run, even though physical reality is extremely complex and the challenges in emulating it remain formidable

Future Work

Physical sensor networks

- Transfer these algorithms to physical multicamera systems and validation
  - *NSF-funded project with the University of California, Riverside*

Better virtual camera sensor networks

- Consistent labeling of pedestrians – Who, What, Where, When?
- Smarter networks through etter cognitive modeling

Bigger, better reality emulators

- Higher-fidelity synthetic video
- Virtual sensor networks in the sky?
- An entire city …

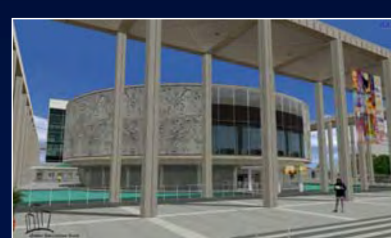
Virtual Los Angeles

Virtual outdoor urban environment

UCLA Urban Simulation Lab

Virtual LA

Building level of detail



Virtual LA

Interior spaces

Acknowledgements

Faisal Qureshi

- PhD thesis developed the Virtual Vision framework

Wei Shao

- PhD thesis developed autonomous pedestrians

Tamer Rabie

- PhD thesis: Active vision in virtual animals/humans

DARPA seedling grant thanks to Tom Strat

- Additional funding from NSF and NSERC Canada

Thank You !

For papers and videos:

cs.ucla.edu/~dt

terzopoulos.com