



# Scalable video architecture

for real time  
tracking and positioning

**Giovanni Gualdi**  
**MMSL, HP Labs, Palo Alto, CA**



# SeViT: (Sentient Video Tracking)

what does it do  
and how

---

# what

SeViT is the video layer to the Sentient sensor networks and provides

- motion detection
- tracking
- positioning
- basic event recognition
- *volumes estimation (not yet)*
- ...

# how

SeViT has to fit the requirements of the sensor space it belongs to:

1. **real time**  
every event triggers an immediate reaction
2. **specific hardware limitations:**
  - no central processing nodes: the processing must be intrinsically distributed
  - no powerful processing nodes
  - no super expansive hardware
3. **easy to deploy**  
“plug&play” + small sizes of the sensors + wireless technologies
4. **scalable and dynamic**  
nodes can be added or removed on the fly
5. **limit the use of the network for video processing**  
other sensors are sharing the same band



SeViT

the hardware

---

# The processing unit P

P is a xScale processor,  
400Mhz, 32MB flash:

- small (~ 7cm x 6cm x 3cm)
- ad-hoc wireless (802.11b wireless compact flash)
- inexpensive (x00\$)
- standard programming platform:
  - Linux operating system
  - libraries (eg. OpenCv) and pre-existing code can be cross-compiled; IPP libraries are available



# The camera



 is a USB web camera:

- inexpensive (x0\$)
- downstream raw or YUV video  
no/low load on the CPU to decompress the video
- USB interface  
no load on the network to downstream the video



# The communication

----- is wireless:

- 802.11b, with 11mbps nominal bandwidth, 4mbps effective
- ad-hoc routing protocol  
central access points avoided





# SeViT

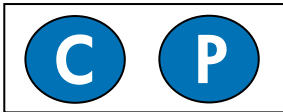
the segmentation node  
the positioning-tracking node

---

# the segmentation node (1)

## *hardware:*

1 Camera + 1 Processing Unit



## *tasks:*

- downstream video from USB
- segmentation
  - background differencing (hysteresis threshold)
- background update
  - median statistic + selectivity
  - single difference for fast ghost suppression
- simple processing for further steps



# the segmentation node (2)

## *considerations:*

- on QVGA (320x240), using 100% of CPU  
~ 6fps
- all the video data is processed in this node
  - no network cast of video streams
  - all the further steps are vision independent
  - optional: down stream the video on the net after being encrypted (3DES)

# the positioning-tracking node (1)

## *hardware:*

1 Processing Unit



## *tasks:*

- positioning
- tracking
  - at 4Hz the most recent data from all the connected segmentation nodes is gathered together and the tracking is computed

# the positioning-tracking node (2)

## *considerations:*

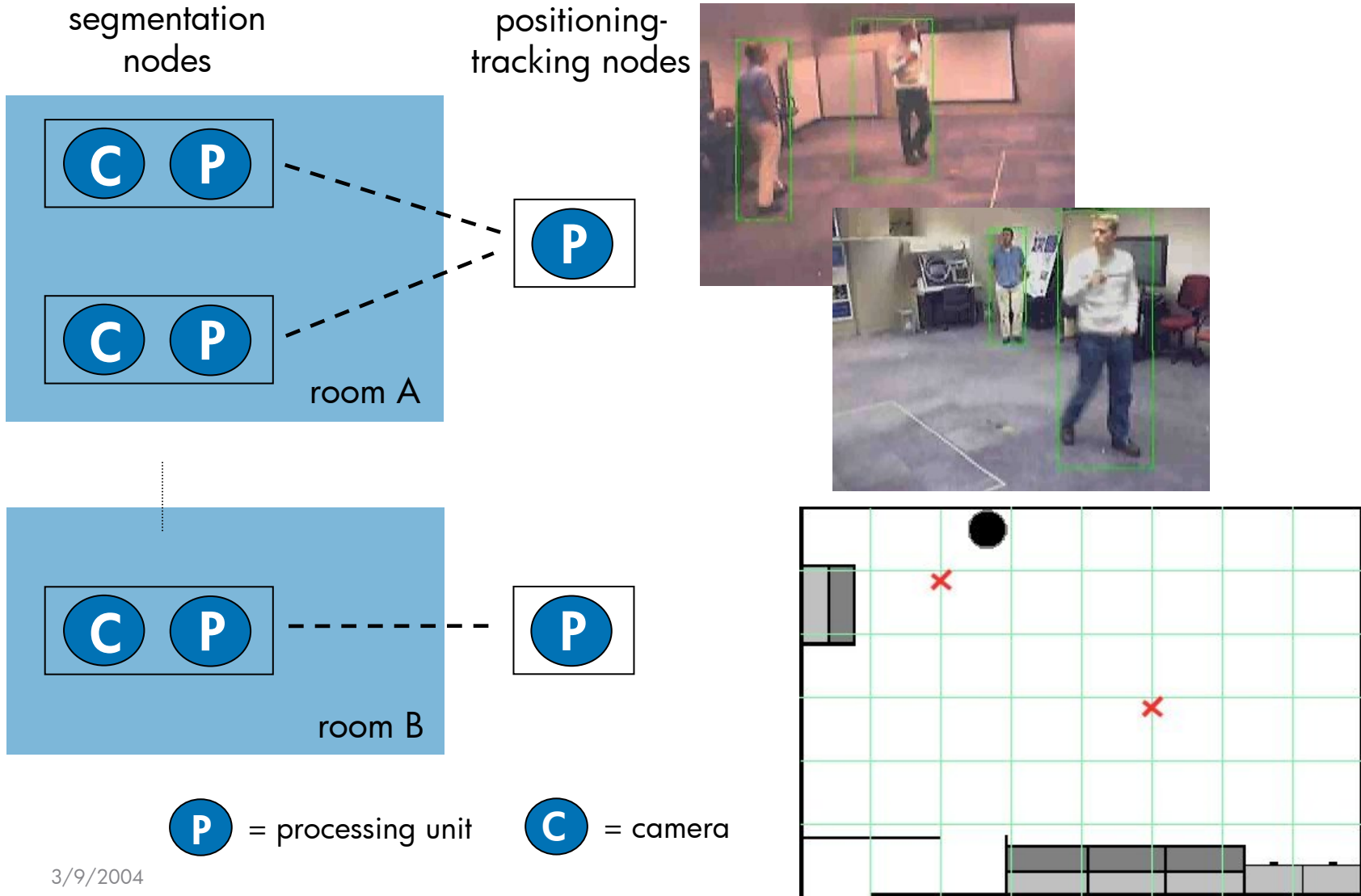
- positioning and tracking work with just 1 camera
  - more cameras will decrease the uncertainties due to occlusions of objects
- no synchronization is required
  - if a segmentation node slows or crashes, the positioning-tracking continues its regular execution
  - segmentation nodes are added/removed on the fly

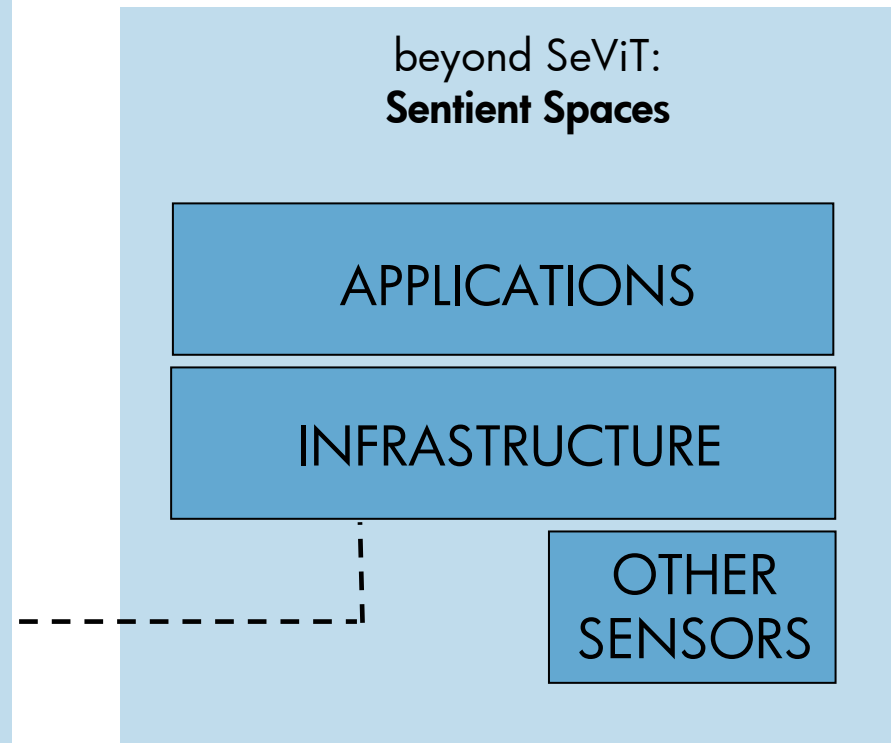
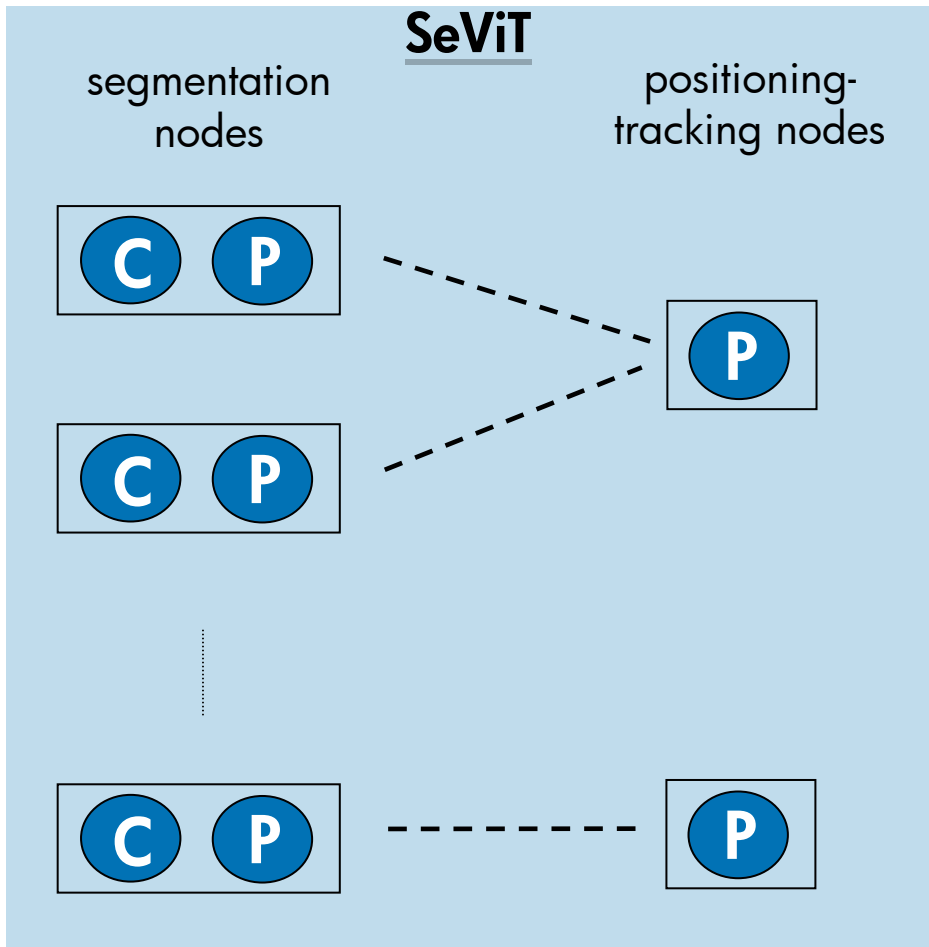
# SeViT

## the architecture

---

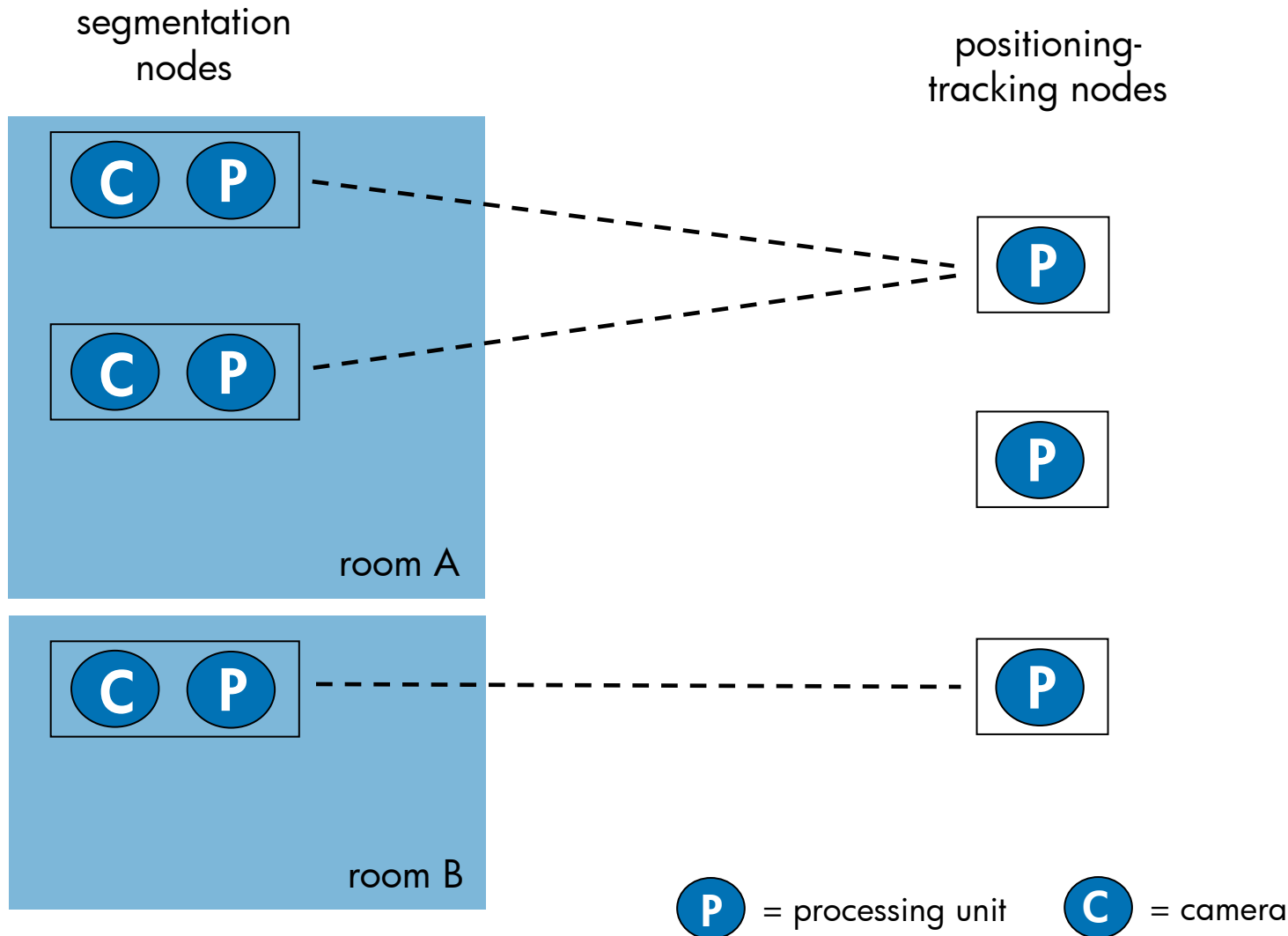
# The structure of the architecture

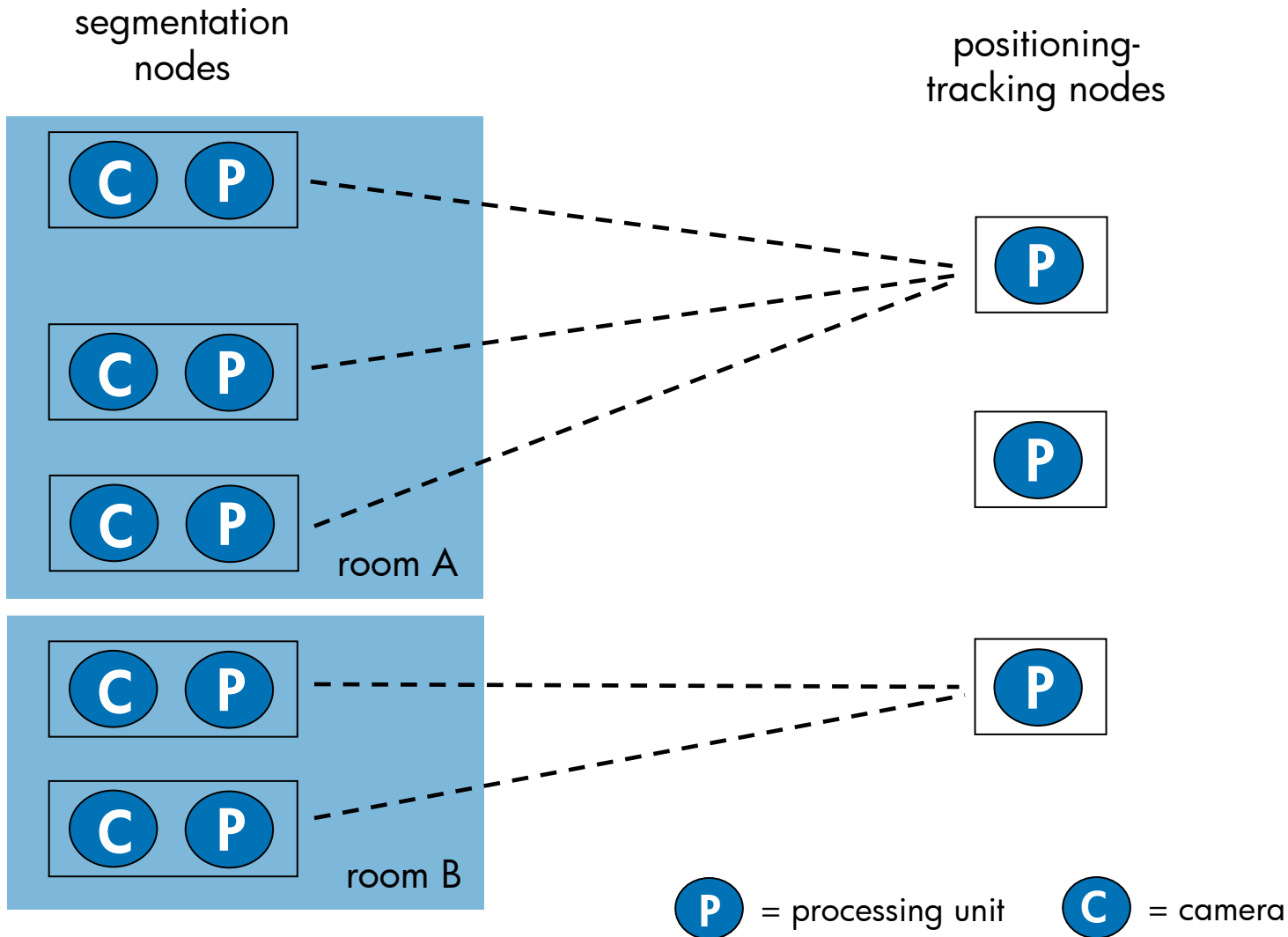






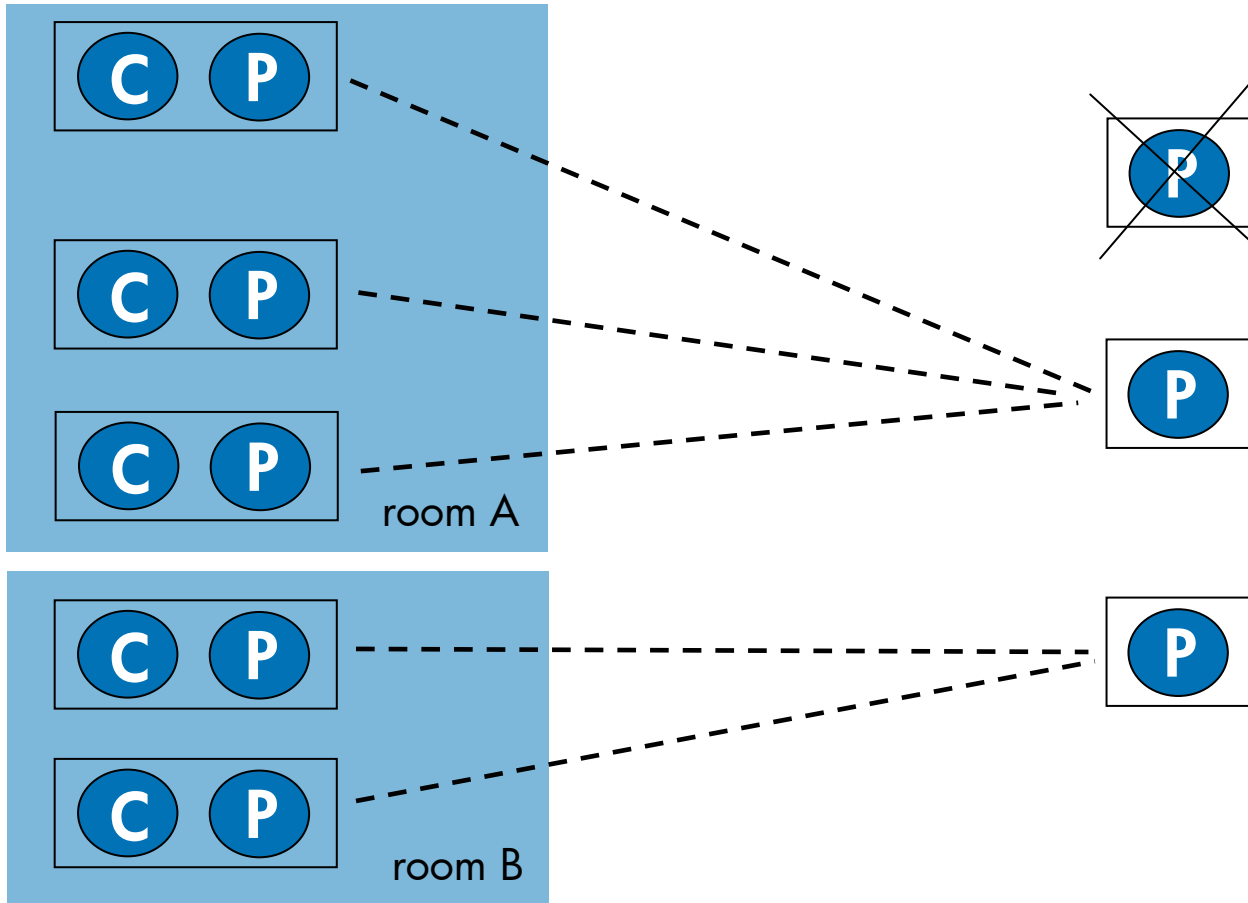
# The structure is scalable and dynamic





segmentation  
nodes

positioning-  
tracking nodes



conclusions

the test bed  
the results

---

# test bed

- indoor room of 28 x 20 feet
- 3 segmentation nodes with 3 different cameras models:
  - 3 cameras at three corner of the room
  - 100% of the room is covered
- 1 PC to display the streaming and the positioning (optional)

# results on the test bed

- **real time**
- **easy to deploy**  
measured deployment time (extrinsic camera calibration + physical deployment + startup of the system): 1 hour
- **scalable and dynamic**
- **inexpensive**  
3 cameras + hardware for tracking and positioning  
~ 1500\$
- **accurate and robust**  
50cm accuracy in positioning



i n v e n t