

# Audio-Visual HRI for Robots in Factories

## Ears and Eyes for Robot



*European Robotics Forum, Odense, 6 March 2012*

*Dr C. Morand, Dr N.M. Robertson*

*Heriot-Watt University*



***Aim: To enhance the robot with intelligent sensing (vision and audio),  
To confer on the robot the ability to perceive and interpret external  
events (combined speech and gesture)***

## **LOCOBOT robot context**

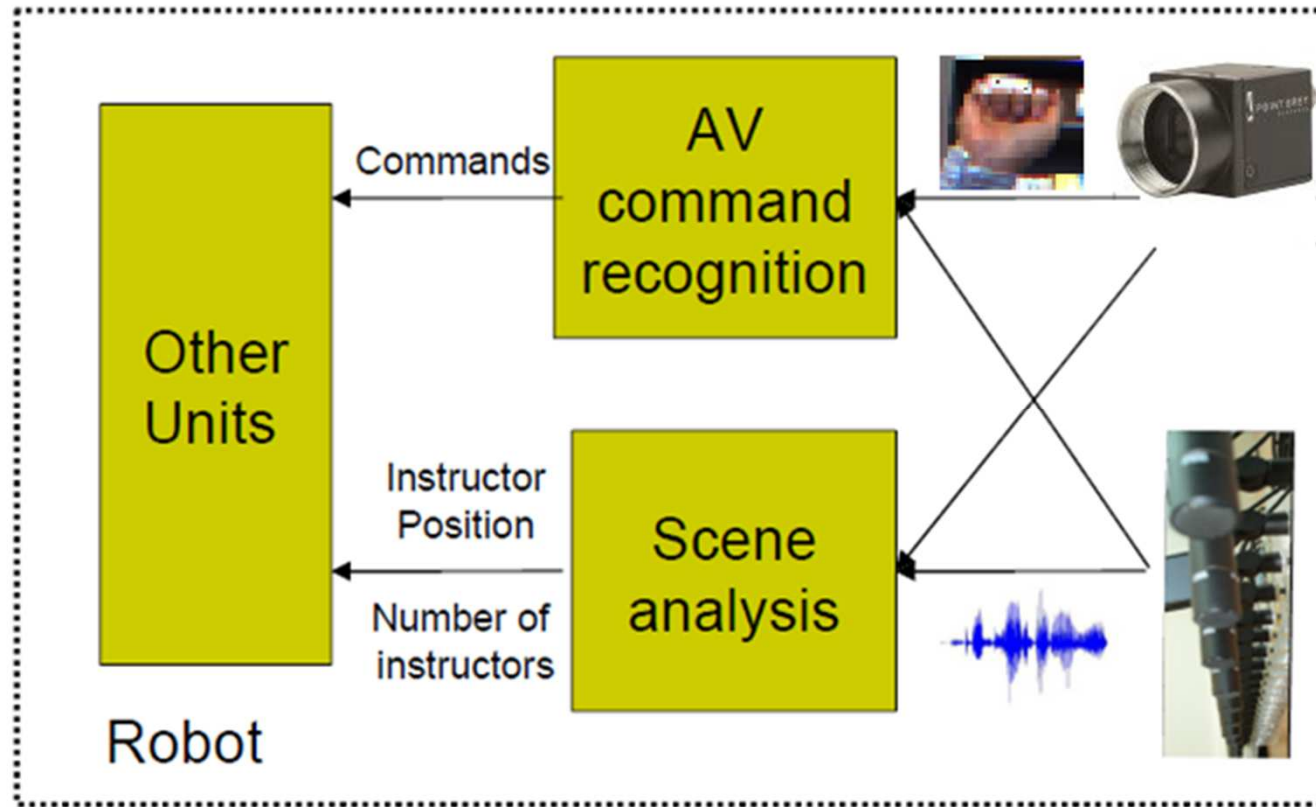
- Co-Worker, mobile robot
- Autonomous execution of tasks
- Unforeseen events managed by Human to Robot Communication
- One designated Human co-worker

## **Beyond state-of-the-art**

- Changing environment, mobile sensors
- SOA: meeting rooms

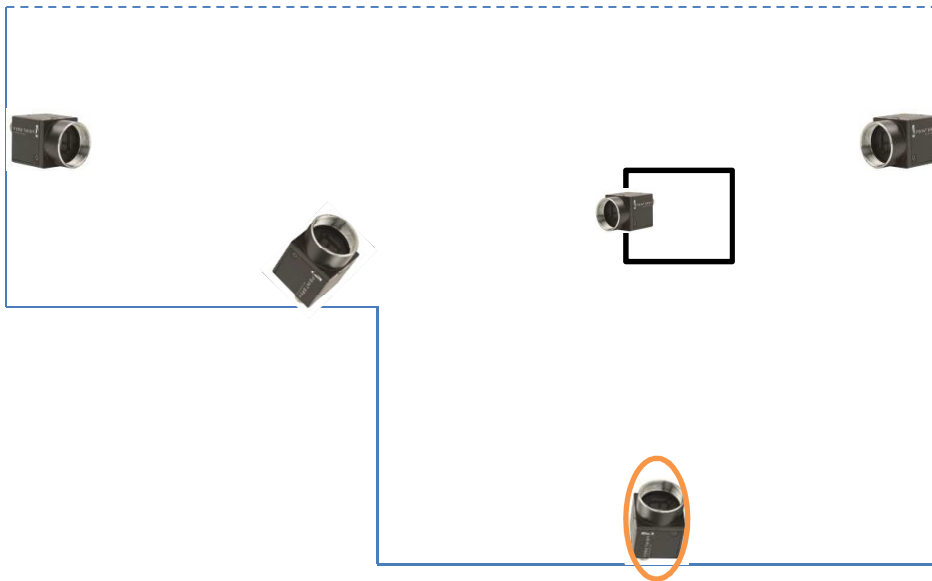


# Audio-Visual Scene Analysis



# Video Sensors set-up

- ☒ Unforeseen event: Continuous gesture monitoring of the human worker
- ☒ Autonomous & mobile robot: independent robot and worker trajectories
- ☒ Continuous people tracking via external and embedded camera network
  - Occlusion in cluttered and changing factory environment : external network
  - Resolution to distinguish between gestures: embedded sensor



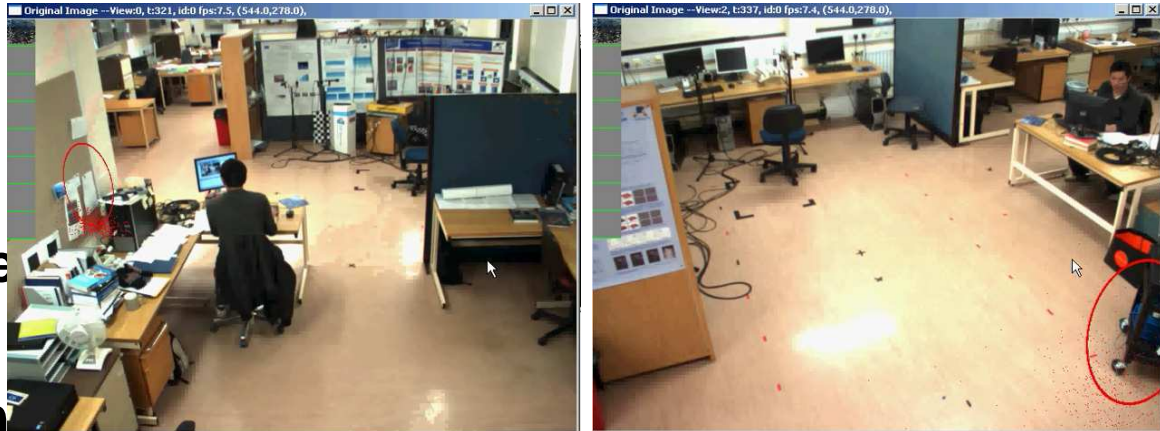
# Video Tracking concept



External network of 4 cameras

# Tracking – External network

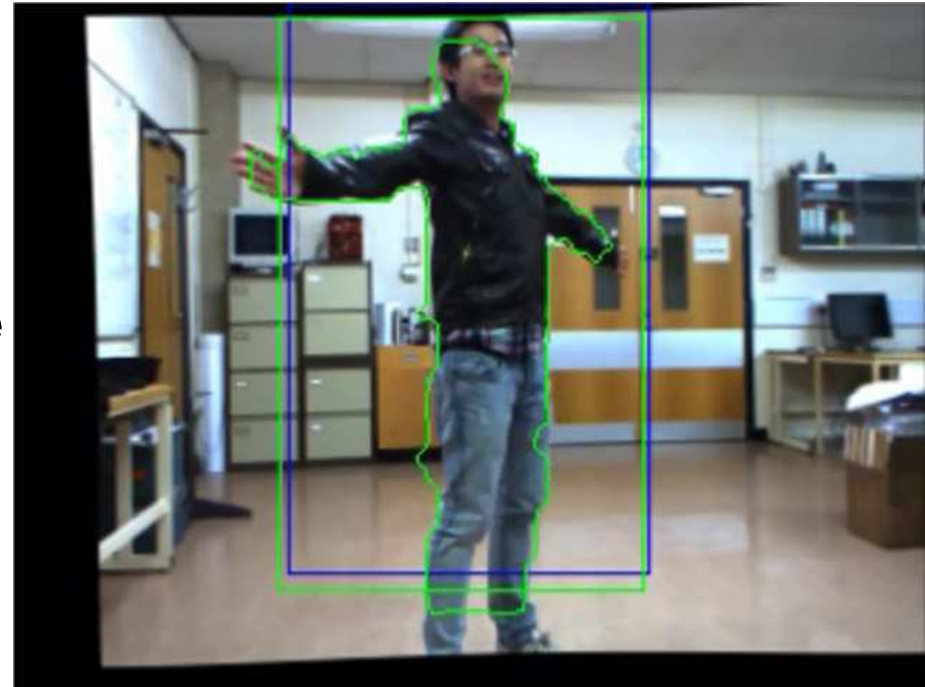
- ☞ Multi-camera, multi-target tracking. Calibrated network.
- ☞ Relative position of the worker and the robot
- ☞ Deals with occlusion
- ☞ Texture feature to characterise people: identification of people
- ☞ Set-up transposable to new location (test in Profactor)



W. Limprasert, A. Wallace, and G. Michaelson. “Accelerated people tracking using texture in a camera network”, VISAPP, Roma, Italy, 24-26 February 2012.

# Tracking – embedded sensor

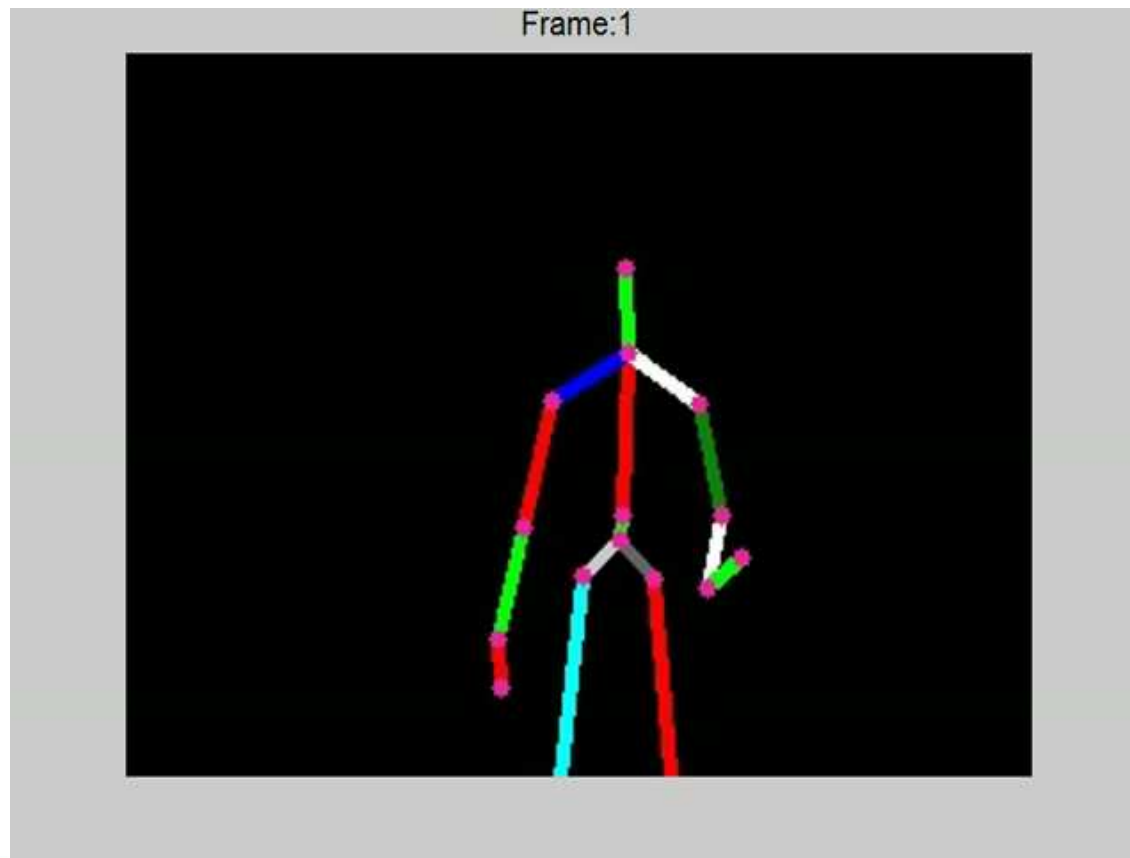
- ☞ Stereo pair for access to depth information
- ☞ Active tracking of designated worker
- ☞ Optimal view for further gesture recognition
- ☞ Precise extraction of contours
- ☞ Plug and play on the robot



W. Gulyanov, C. Morand, N.M. Robertson, and A.M. Wallace. “Real-time active visual tracking with level sets”. Proc. 4th International Conference on Imaging for Crime Detection and Prevention ICDP-11, 2011.

# Gesture recognition

 **Concept verified using HMM and sliding window**





## Tracking

- Occlusion.
- Embedded mobile sensors
- Recovery of errors.

## Gesture recognition

- Monitoring of gesture
  - Begin / end
  - Gesture/non/gestures
- View invariant gesture recognition
- Invariant to duration
- Robustness to intra and inter human variability

## Toward a resolution of the problems by combining with audio...

# Thanks for your attention!

---

## Contact (HWU)

Dr C.Morand

Heriot-Watt University  
Edinburgh, UK

+44 131 451 3325

[C.Morand@hw.ac.uk](mailto:C.Morand@hw.ac.uk)