TASK-DEPENDENT SALIENCY ESTIMATION FROM TRAJECTORIES OF AGENTS IN VIDEO SEQUENCES
Damian Campo, Mohamad Baydoun, Lucio Marcenaro, Carlo Regazzoni.

University of Genoa DITEN

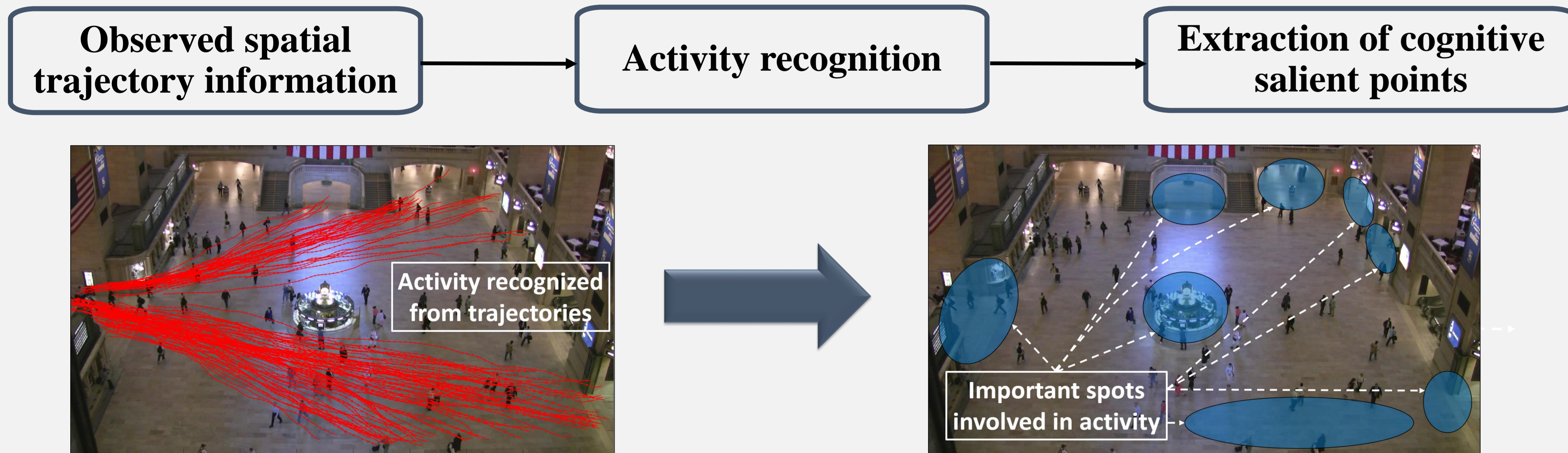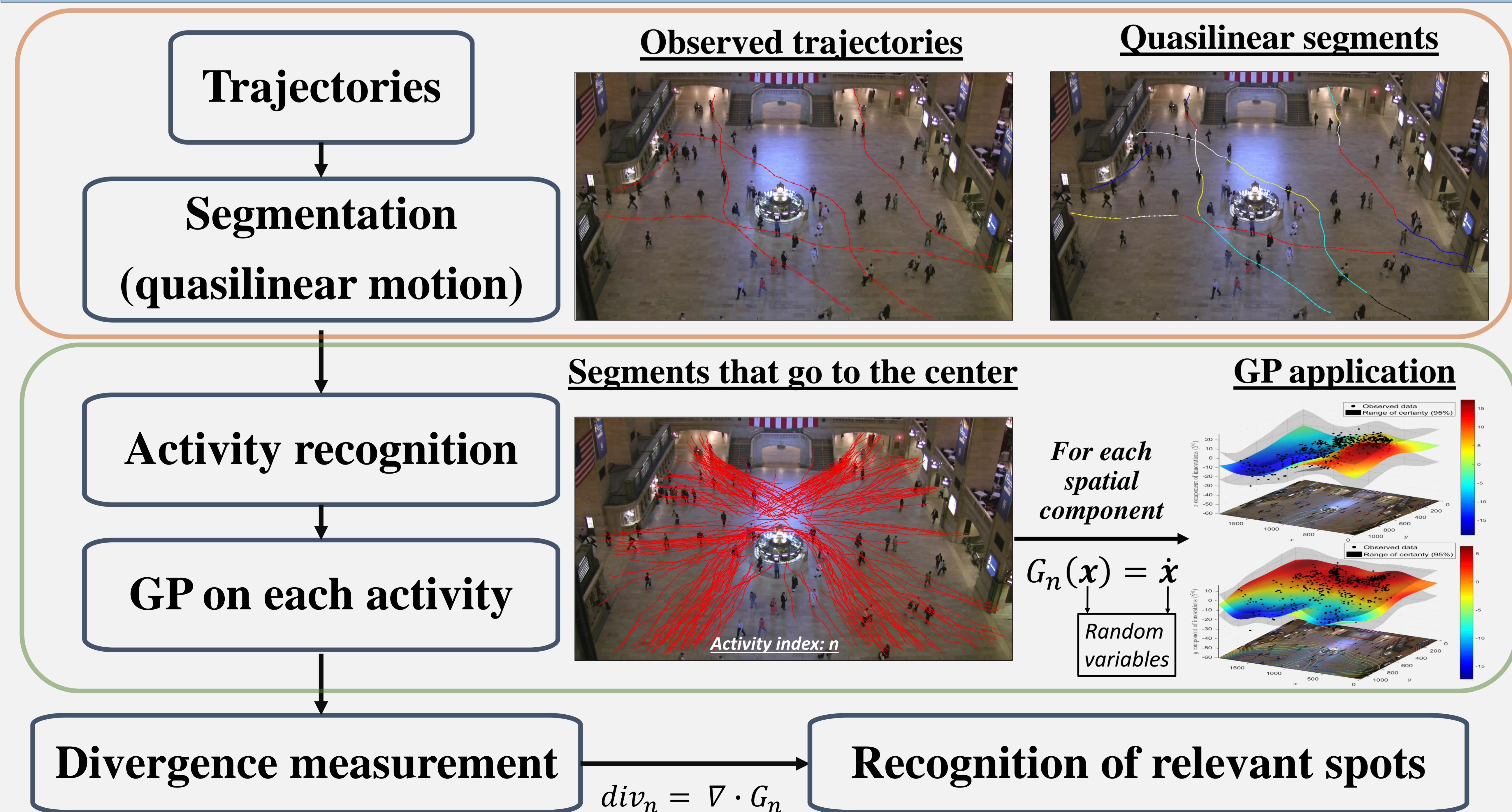Universidad Carlos III de Madrid

INTRODUCTION

Traditional saliency approaches consider only image features such as color or shape for identifying important parts of the scene (bottom-up approach) [1]. Most of works [2,3] are based on such approach and do not consider task dependent cues for estimating relevant zones in scenes (top-down approach) [4]. This work analyzes **cognitive behaviors** of moving entities to **estimate areas of interest** in video scenes in a **top-down fashion**.


Observed spatial trajectory information → Activity recognition → Extraction of cognitive salient points

Activity recognized from trajectories → Important spots involved in activity

PROPOSED METHOD

Observed trajectories

Quasilinear segments

Trajectories → **Segmentation (quasilinear motion)**

Segments that go to the center

Activity index: n

GP application

For each spatial component

$G_n(x) = \dot{x}$

Random variables

Observed data
Range of certainty (95%)

Activity recognition → **GP on each activity**

Divergence measurement

$div_n = \nabla \cdot G_n$

Recognition of relevant spots

CASES OF STUDY

Task dependent salient points in scenarios (ground truth)

Pedestrian dataset
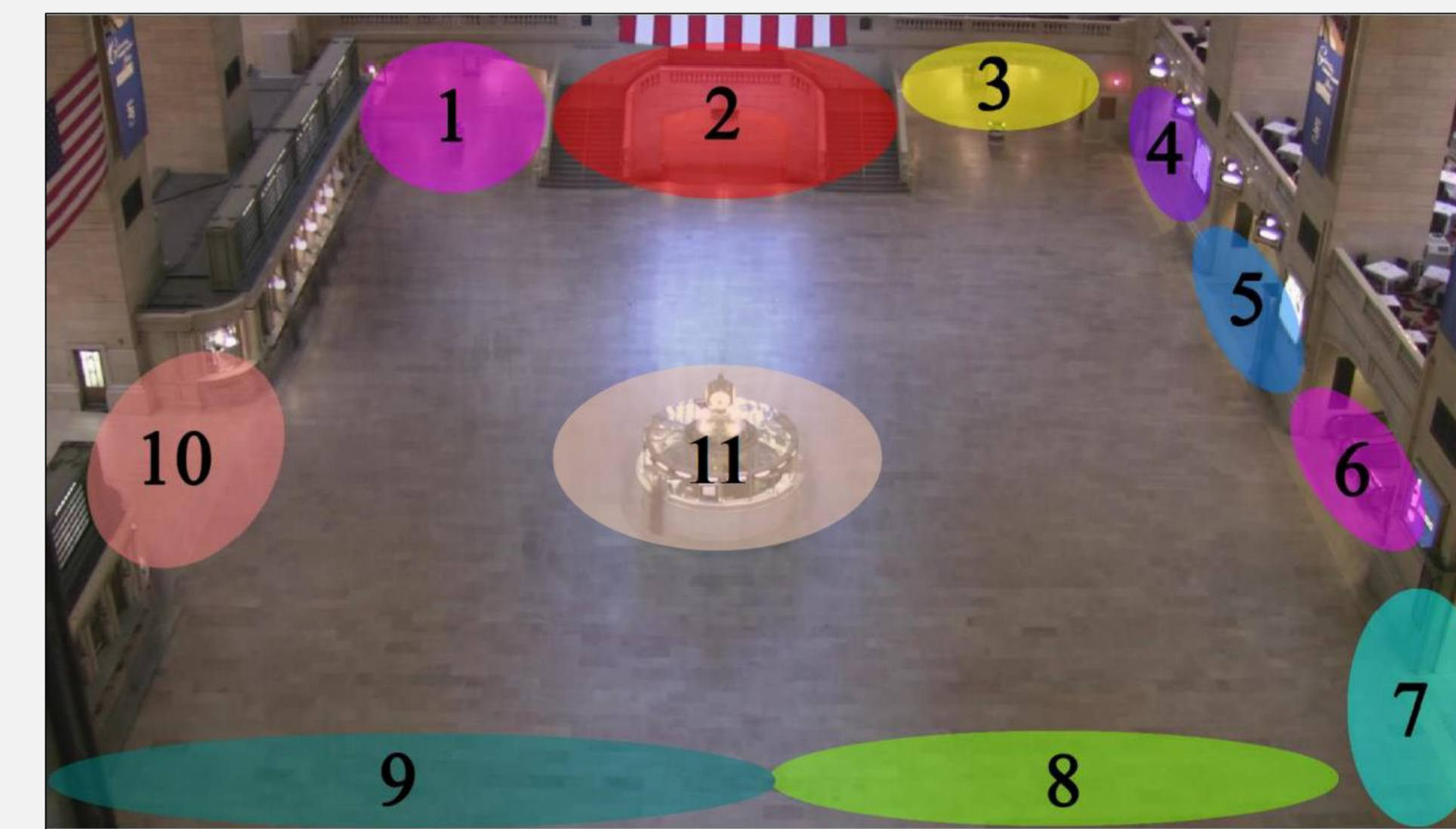
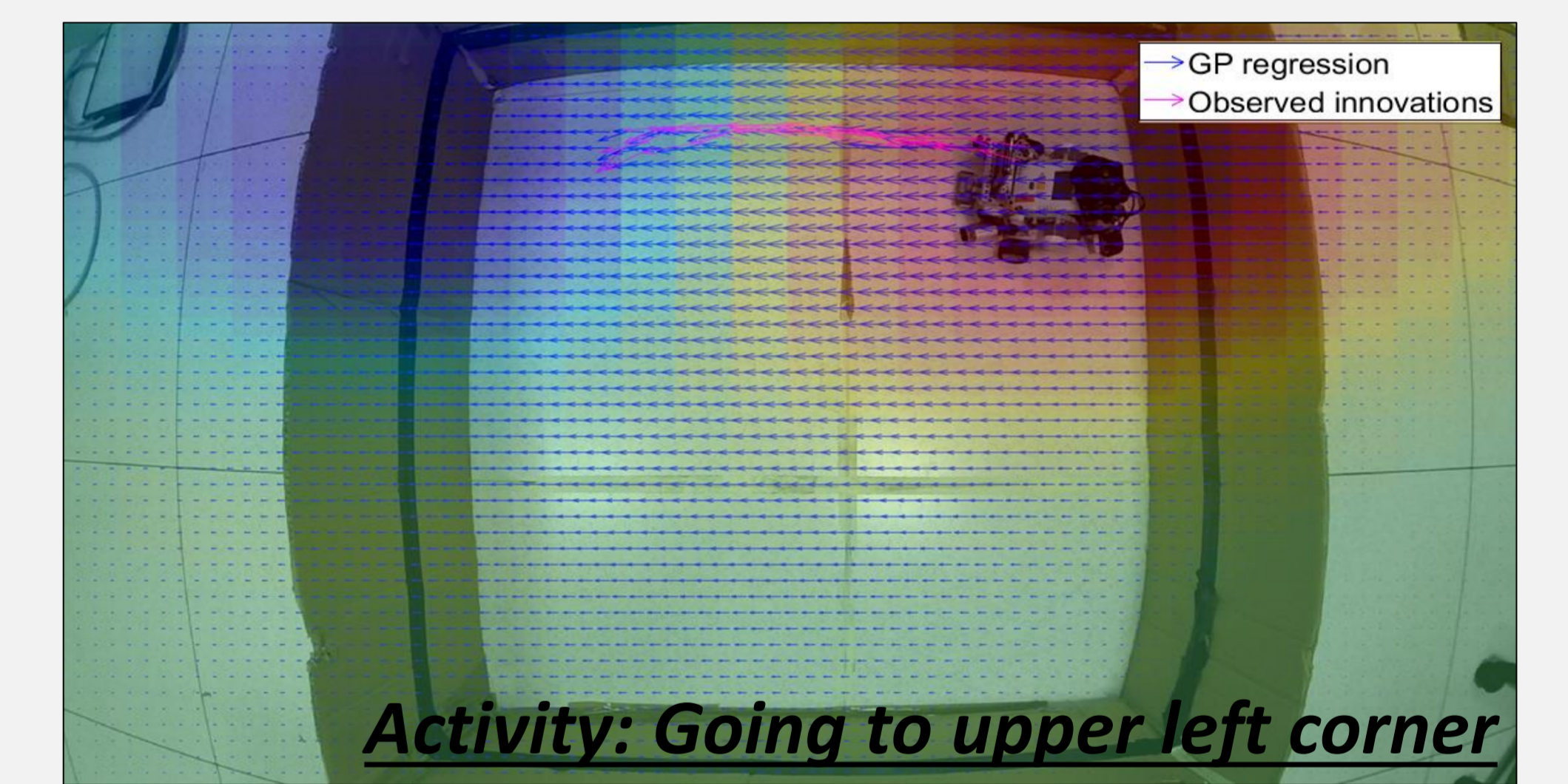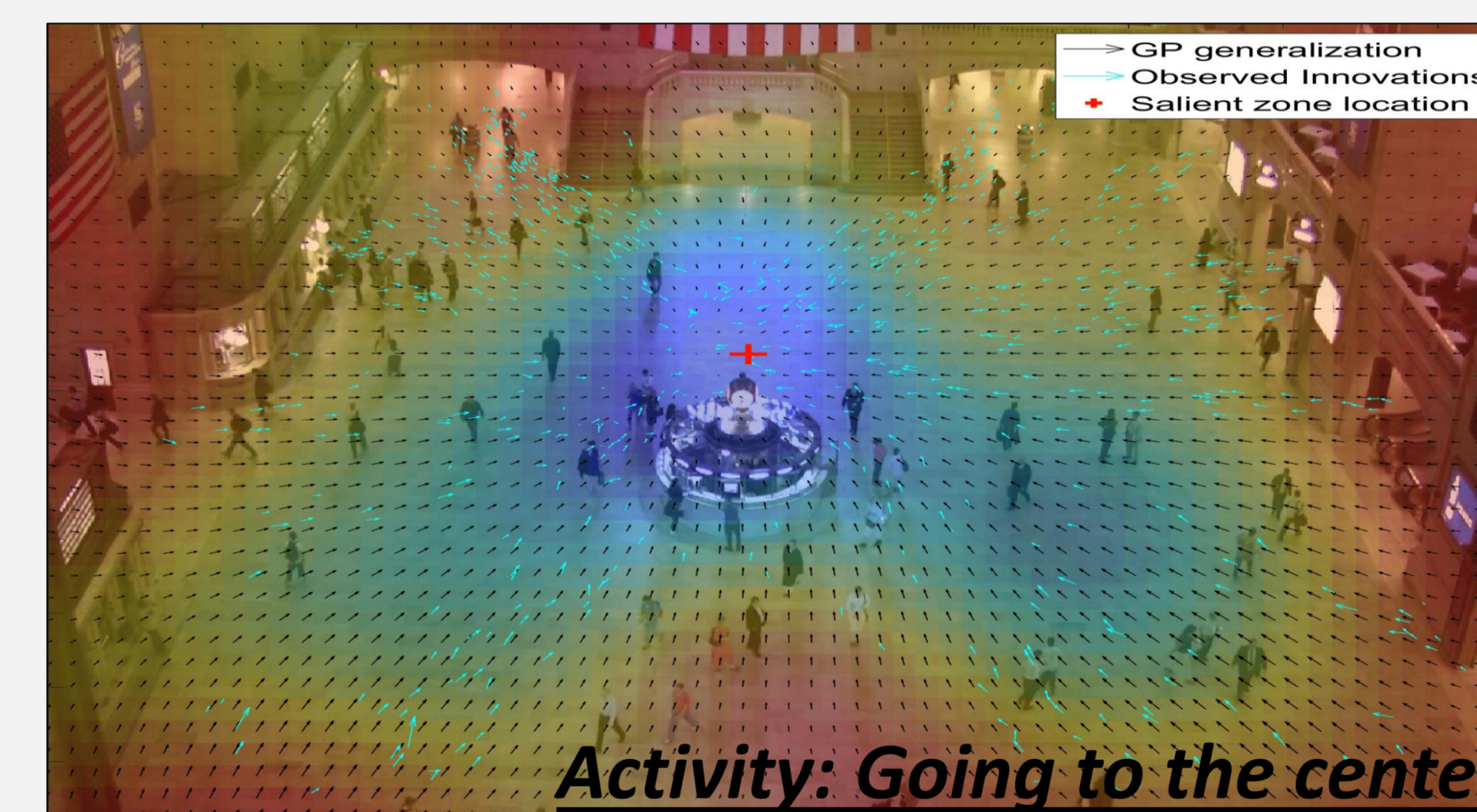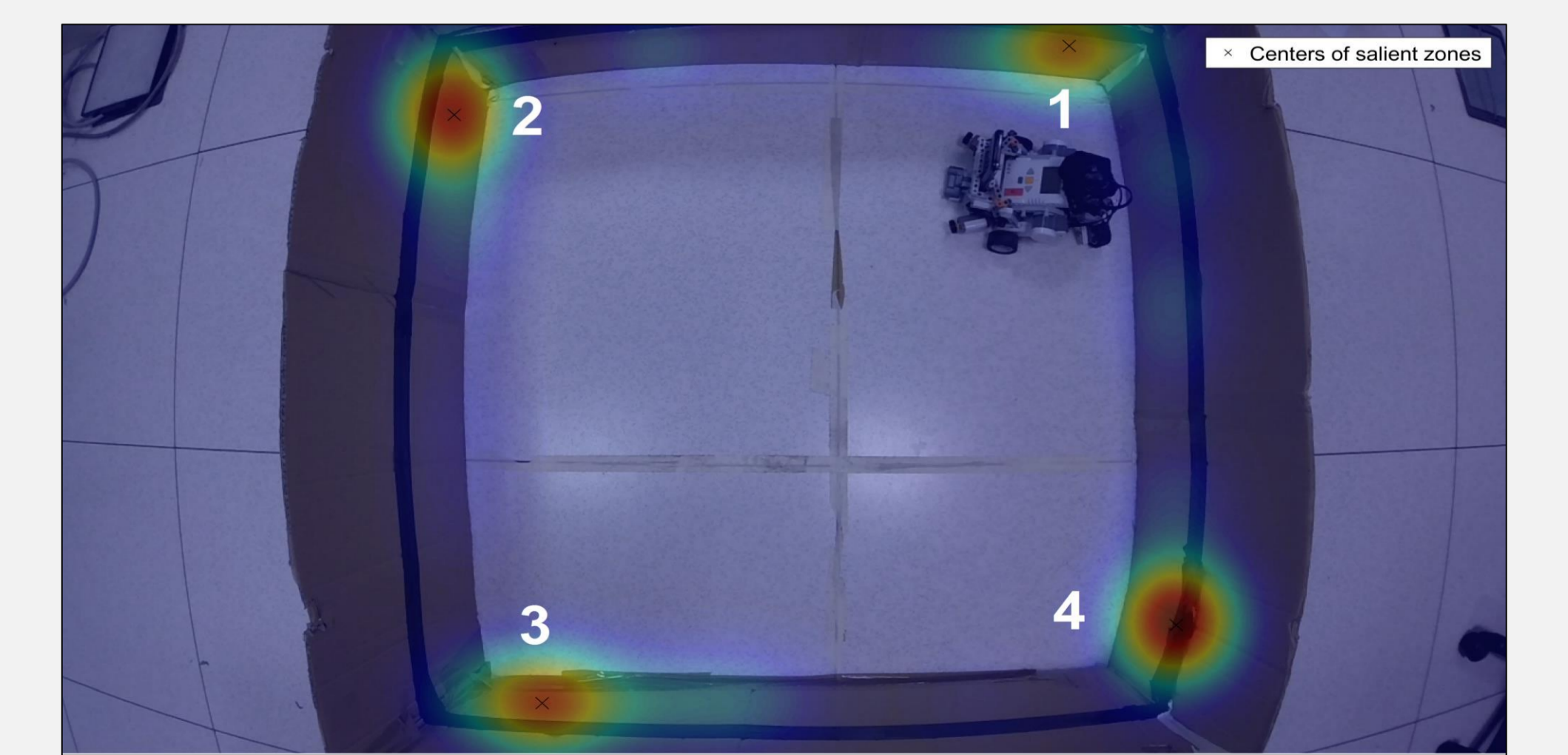**11 areas** identified as **entries, exits** and **obstacles** in the environment

Moving robot Dataset
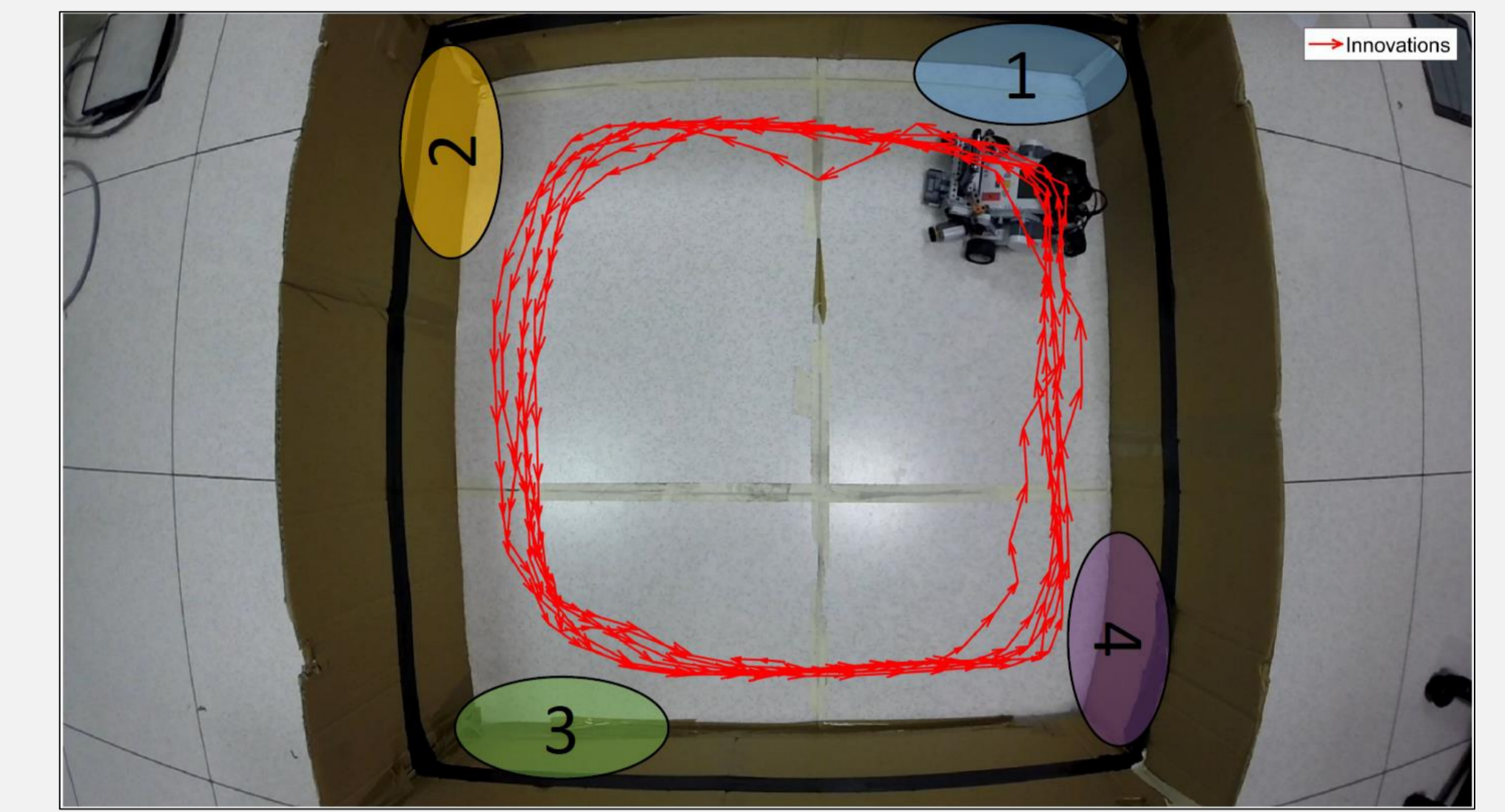
4 areas identified as **attractive points** for a control perimeter task

RESULTS

Divergence maps for single activities

GP generalization
Observed Innovations
Salient zone location

Activity: Going to the center

GP regression
Observed innovations

Activity: Going to upper left corner

Identified task depending salient points

Merged centers of force

Centers of salient zones

References

[1] L. Itti and C. Koch, "Computational modelling of visual attention," Nature Reviews Neuroscience, vol. 2, no. 3, pp. 194–203, 2001.

[2] Li, F. Meng, and J. Mao, "Saliency detection on videos with scene change," Audio, Language and Image Processing (ICALIP) 2015, pp. 506–510.

[3] O. Le Meur, P. Le Callet, and D. Barba, "Predicting visual fixations on video based on low-level visual features," Vision Research, vol. 47, no. 19, pp.2483–2498, 2007.

[4] D.H. Ballard and M.M. Hayhoe, "Modelling the role of task in the control of gaze," Visual Cognition, vol. 17, no. 6-7, pp. 1185–1204, 2009.