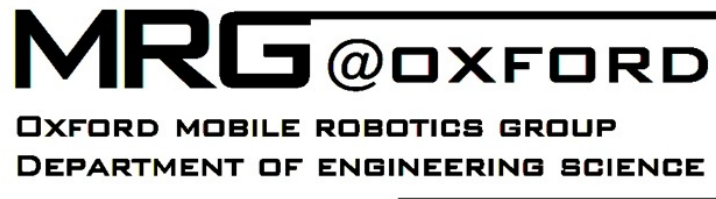


Shady Dealings

Techniques for Long-term Outdoor Visual Localisation

Alex Stewart

Mobile Robotics Group, Oxford University

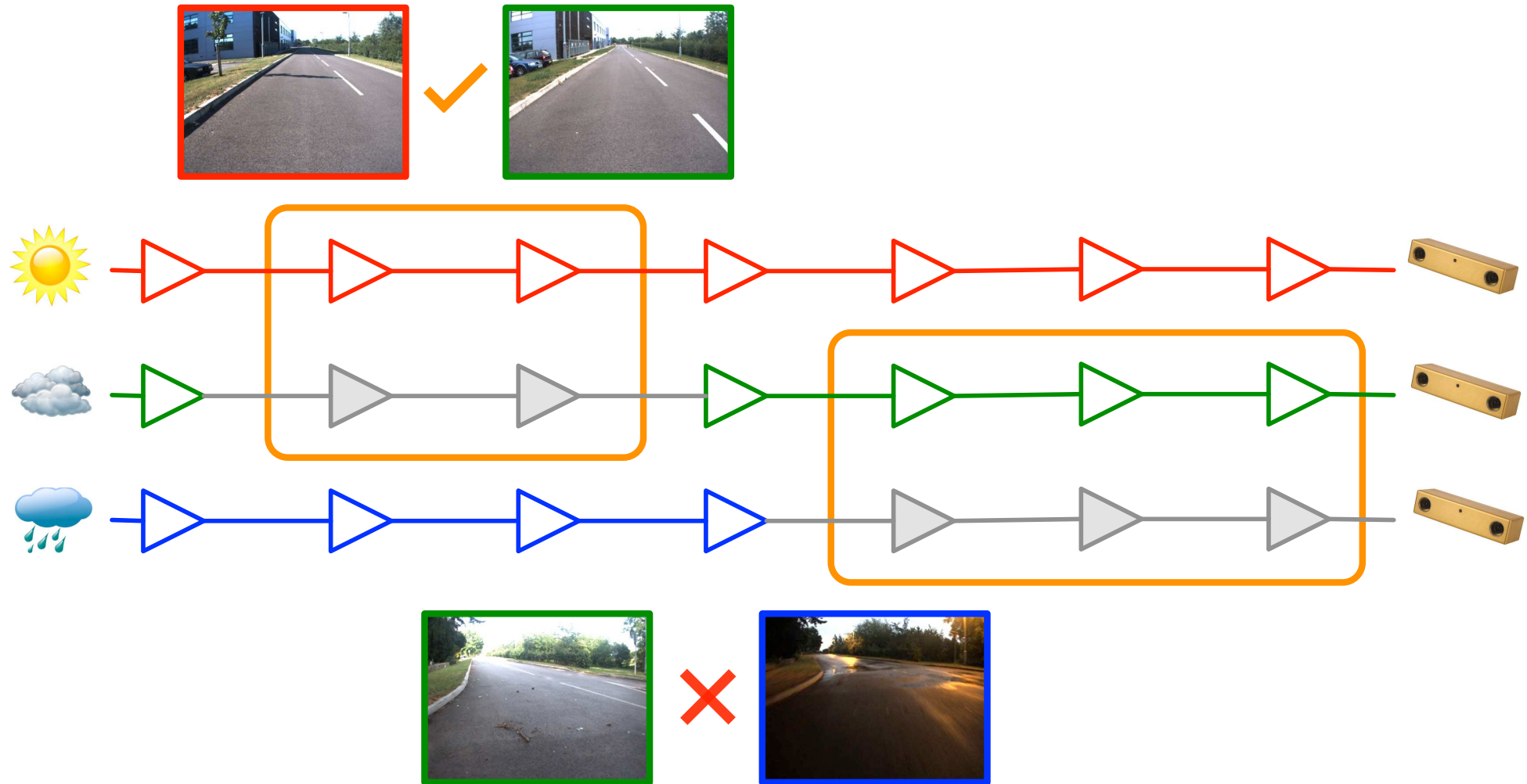


How do we navigate in changing environments?



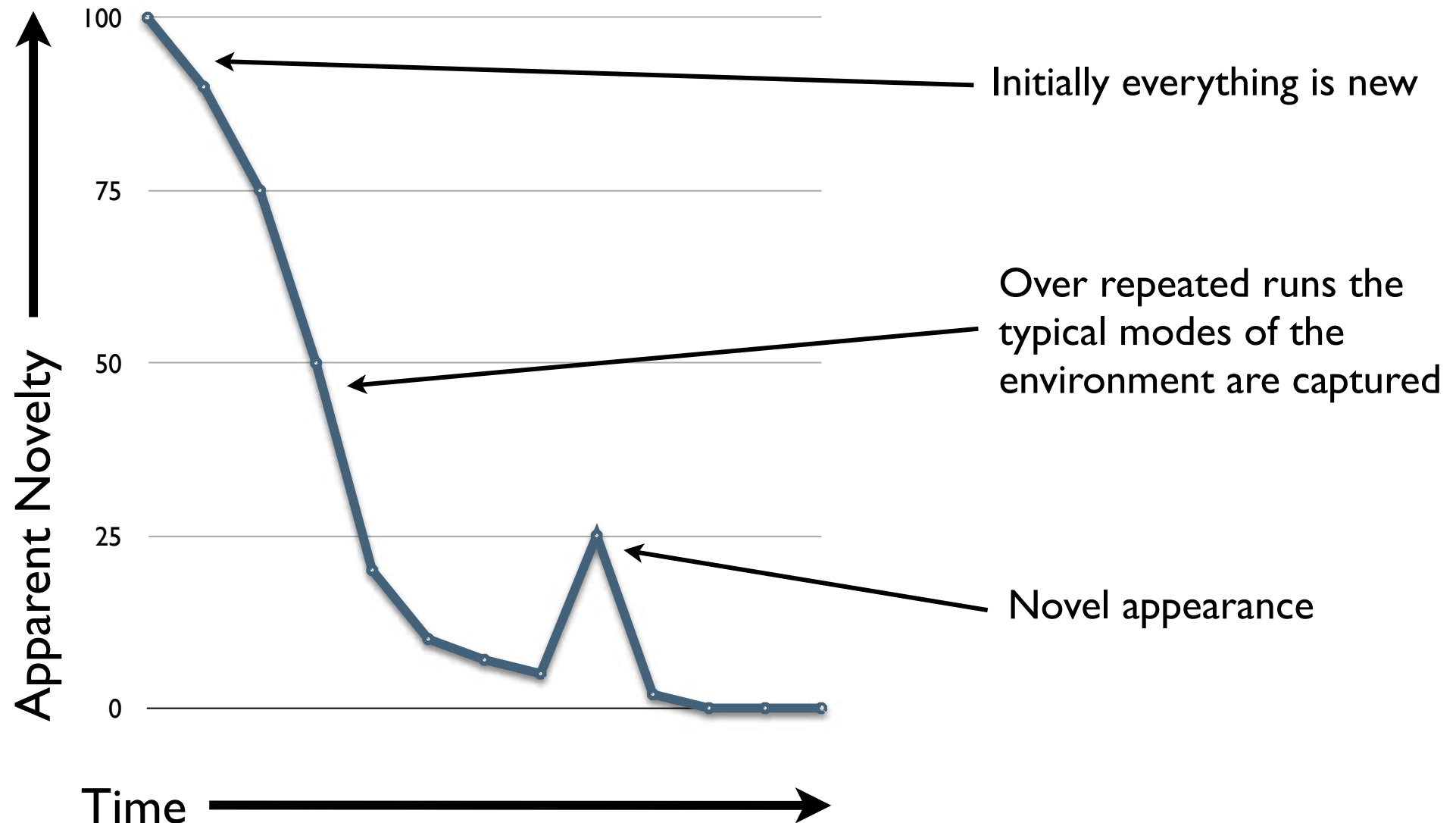
- Appearance changes with (*at least*):
 - Time of day & year, weather and dynamic objects.
- Why should one representation capture all of this?
⇒ Experience Based Navigation

Churchill, W., & Newman, P., “*Experience-based navigation for long-term localisation*”,
The International Journal of Robotics Research (IJRR), 32(14), 1645-1661.

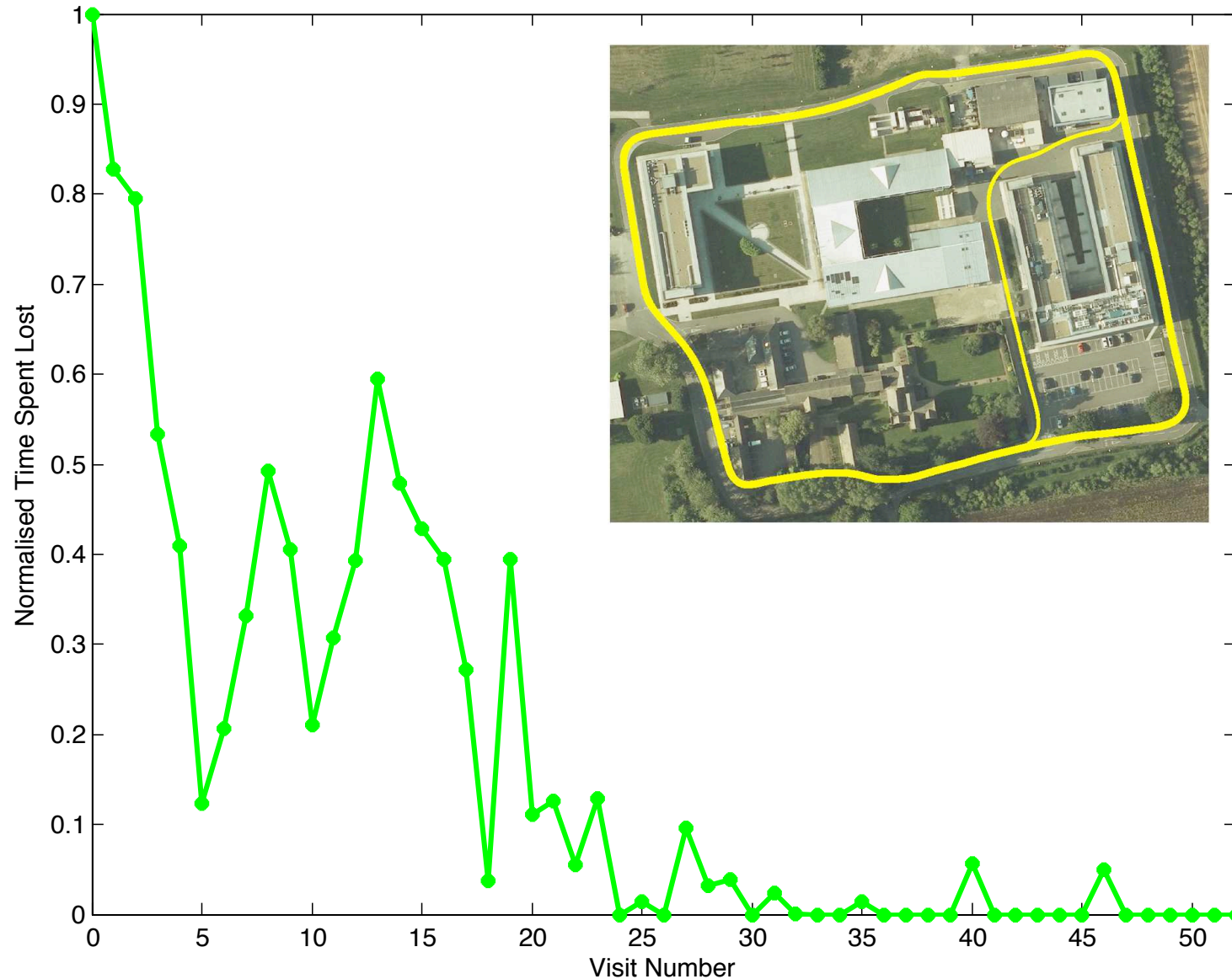


- Always run VO, opportunistically match to Experiences
- Save new Experience only when your current Experiences are insufficient

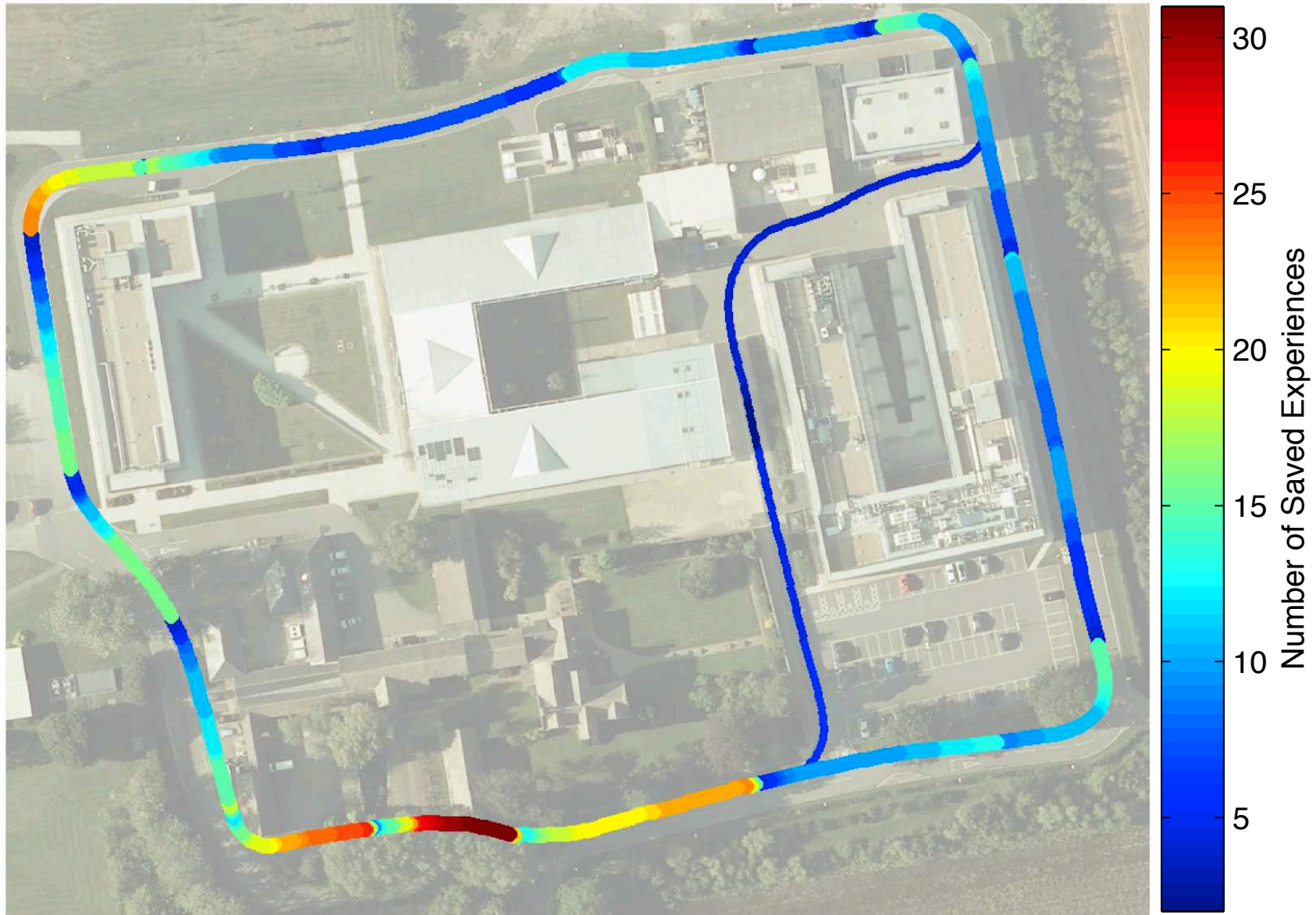
How many Experiences - what should we expect?



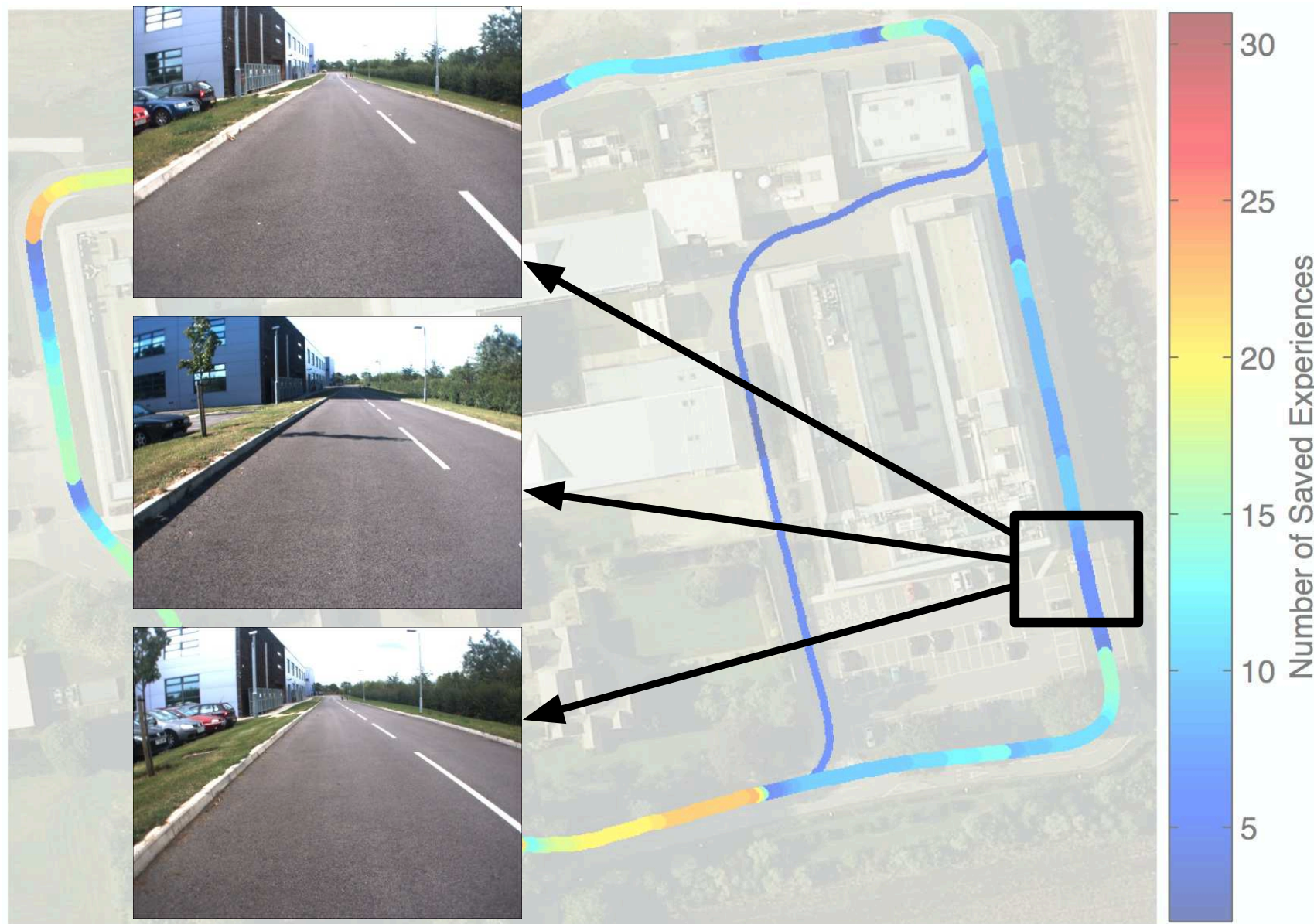
Real Data - 3 Months, 37km, 53 loops



Where did we save Experiences?



Consistent area - few Experiences required



(Dynamic) Shadows - more Experiences



Can we remove illumination changes?



9:00



17:00

- Is there a colour space which depends on the material properties, not illumination?
 - RGB to Gray is just a linear colour space transform
- Such a transform would *not* be scene specific

Maddern W. et al, "Illumination Invariant Imaging: Applications in Robust Vision-based Localisation, Mapping and Classification for Autonomous Vehicles", ICRA 2014

Can we remove illumination changes? - Yes!

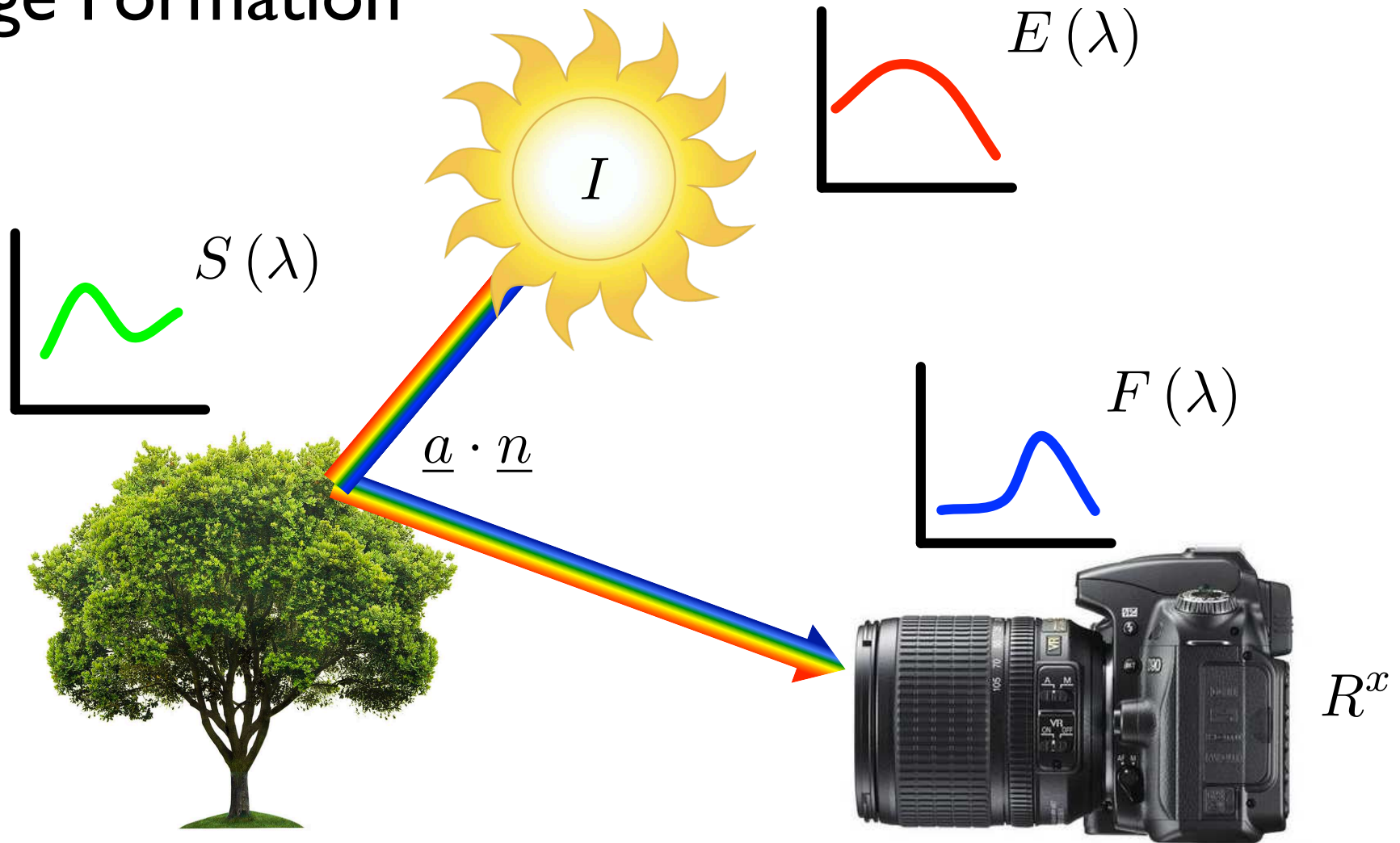


9:00



17:00

Image Formation

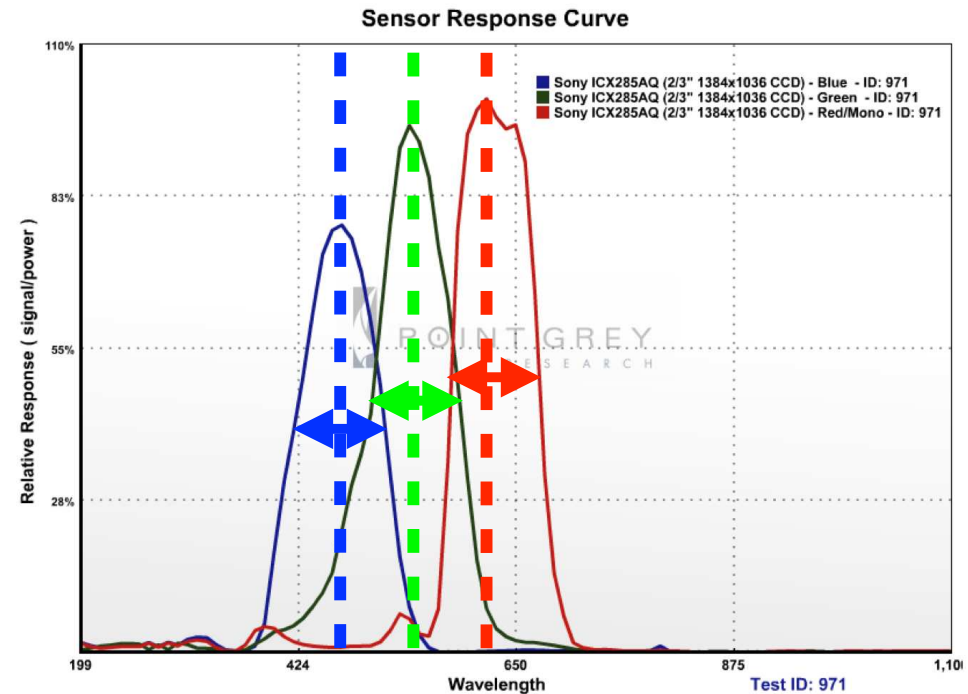


$$R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x \int \underline{S}^x(\lambda) \underline{E}^x(\lambda) \underline{F}(\lambda) d\lambda$$

Assumption #1 - Dirac-delta sensor response



$$F(\lambda) \approx \delta(\lambda_i)$$



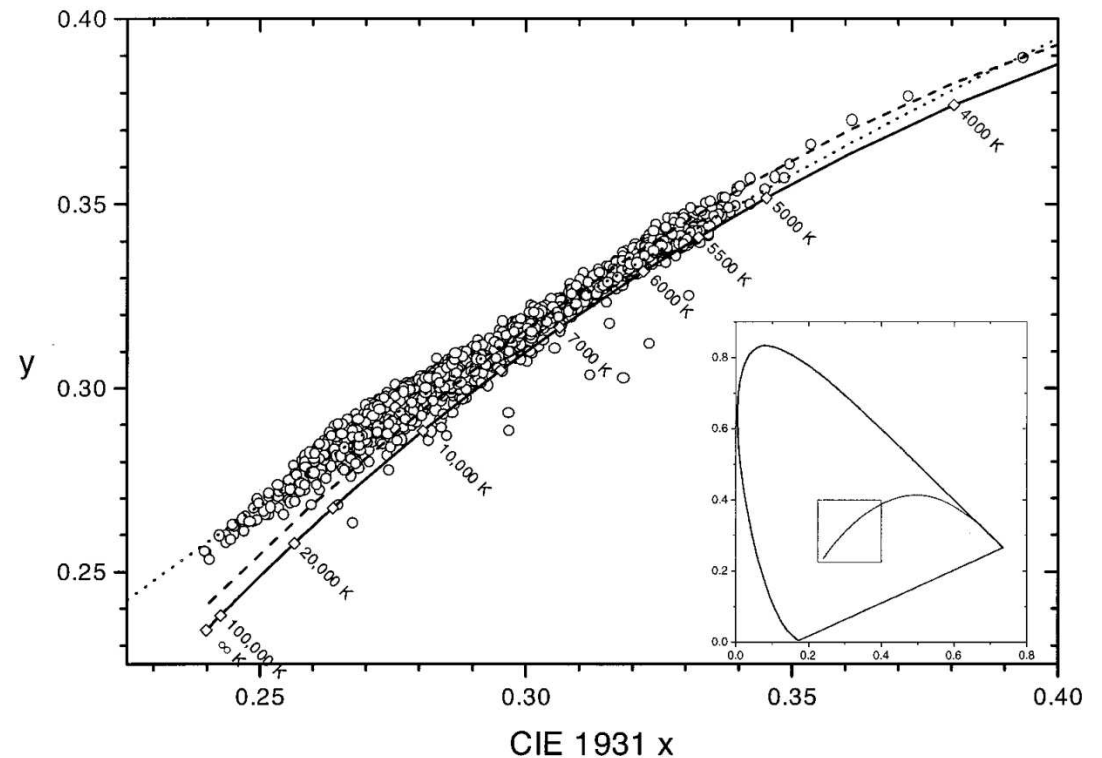
$$R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x \int \underline{S}^x(\lambda) \underline{E}^x(\lambda) \underline{F}(\lambda) d\lambda$$

$$\Rightarrow R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x \underline{S}^x(\lambda_i) \underline{E}^x(\lambda_i)$$

Assumption #2 - Black-body Illumination



$$E(\lambda, T) = 2hc^2 \lambda^{-5} \exp\left(\frac{-hc}{k_b T \lambda}\right)$$



2600 spectra in all weather conditions
over 2 years in Granada, Spain

$$R^{x,E} = \underline{a}^x \cdot \underline{n}^x I^x \underline{S}^x(\lambda_i) \underline{E}^x(\lambda_i) \quad \#1$$

$$\log(R^{x,E}) = \log(\underline{a}^x \cdot \underline{n}^x I^x) + \log(\underline{S}^x(\lambda_i)) + \log(\underline{E}^x(\lambda_i))$$

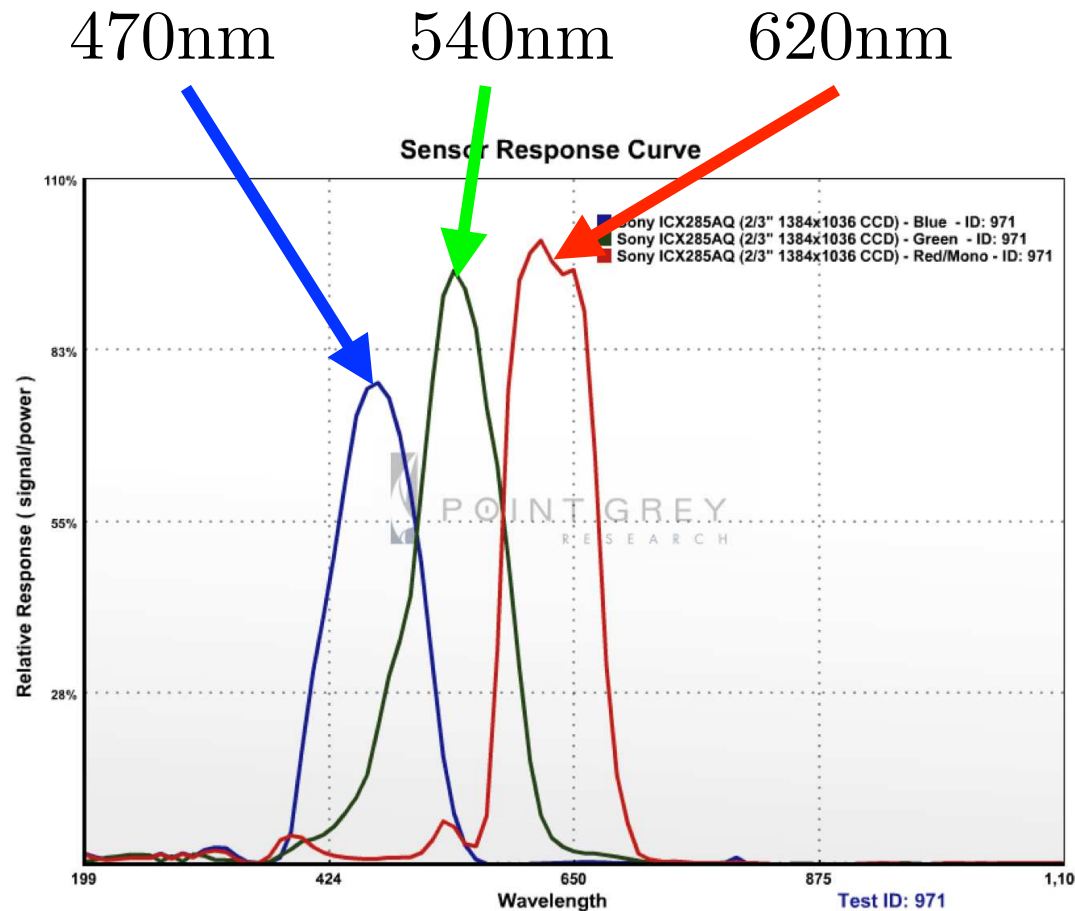
$$\log(R_i) = \log(\underline{a} \cdot \underline{n} I) + \log(\underline{S}_i \cdot f(\lambda_i)) - \underline{g}(T, \lambda_i) \quad \#2$$

Illumination Invariant Image Transform*

$$\mathcal{I} = \log(R_2) - \alpha \log(R_1) - (1 - \alpha) \log(R_3)$$

$$\frac{1}{\lambda_2} = \frac{\alpha}{\lambda_1} + \frac{(1 - \alpha)}{\lambda_3} \quad \lambda_1 < \lambda_2 < \lambda_3$$

* S. Ratnasingam and S. Collins, "Study of the photodetector characteristics of a camera for color constancy in natural scenes", J. Opt. Soc. Am. A 27, 286-294 (2010).



$$\frac{1}{\lambda_2} = \frac{\alpha}{\lambda_1} + \frac{(1 - \alpha)}{\lambda_3}$$

$$\frac{1}{540} = \frac{\alpha}{470} + \frac{(1 - \alpha)}{620}$$

$$\alpha = 0.4642$$

- Obtain λ_i directly from camera specsheet
- **One** line in MATLAB - try it!

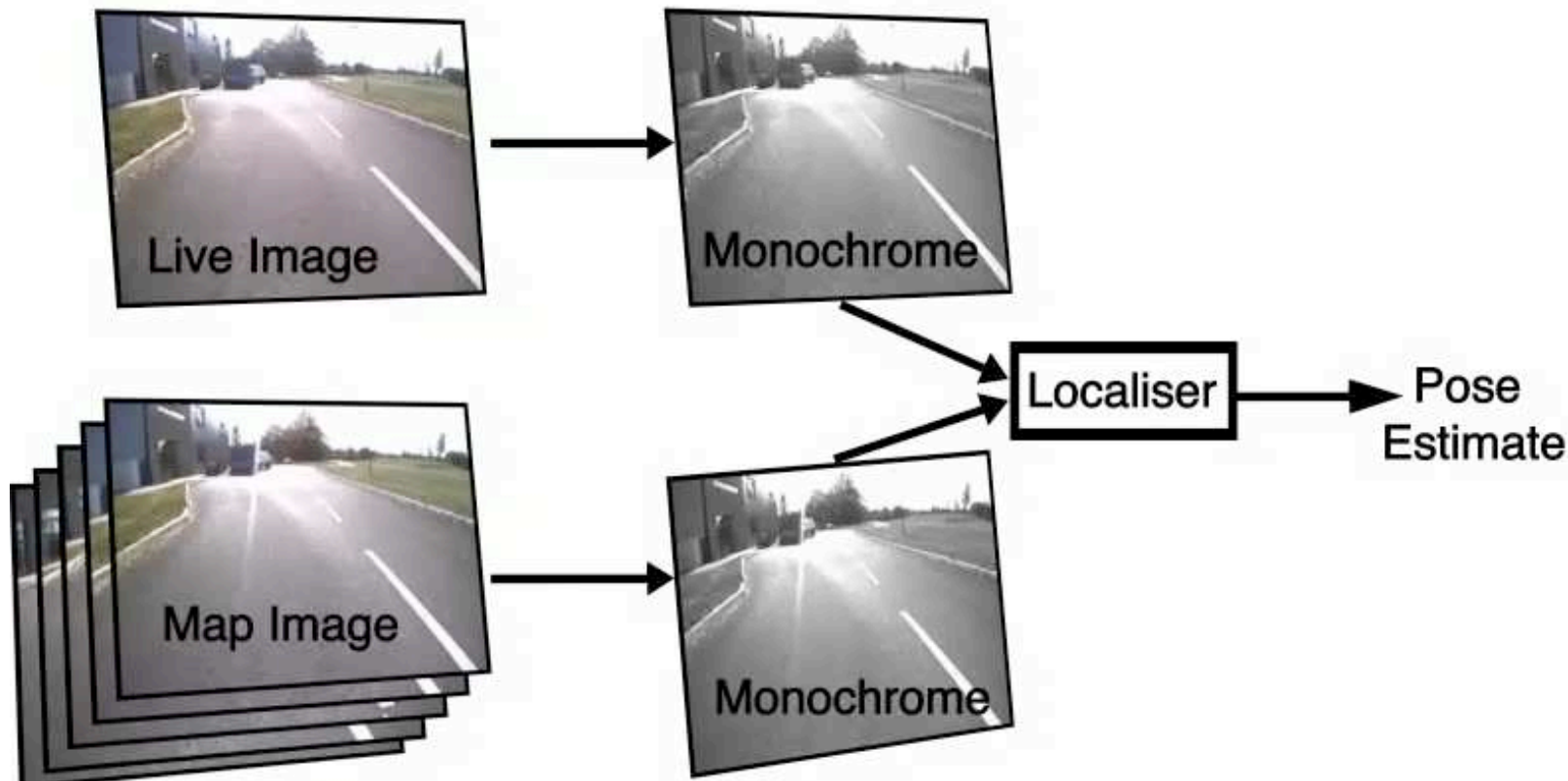
Performance on 24hr Dataset



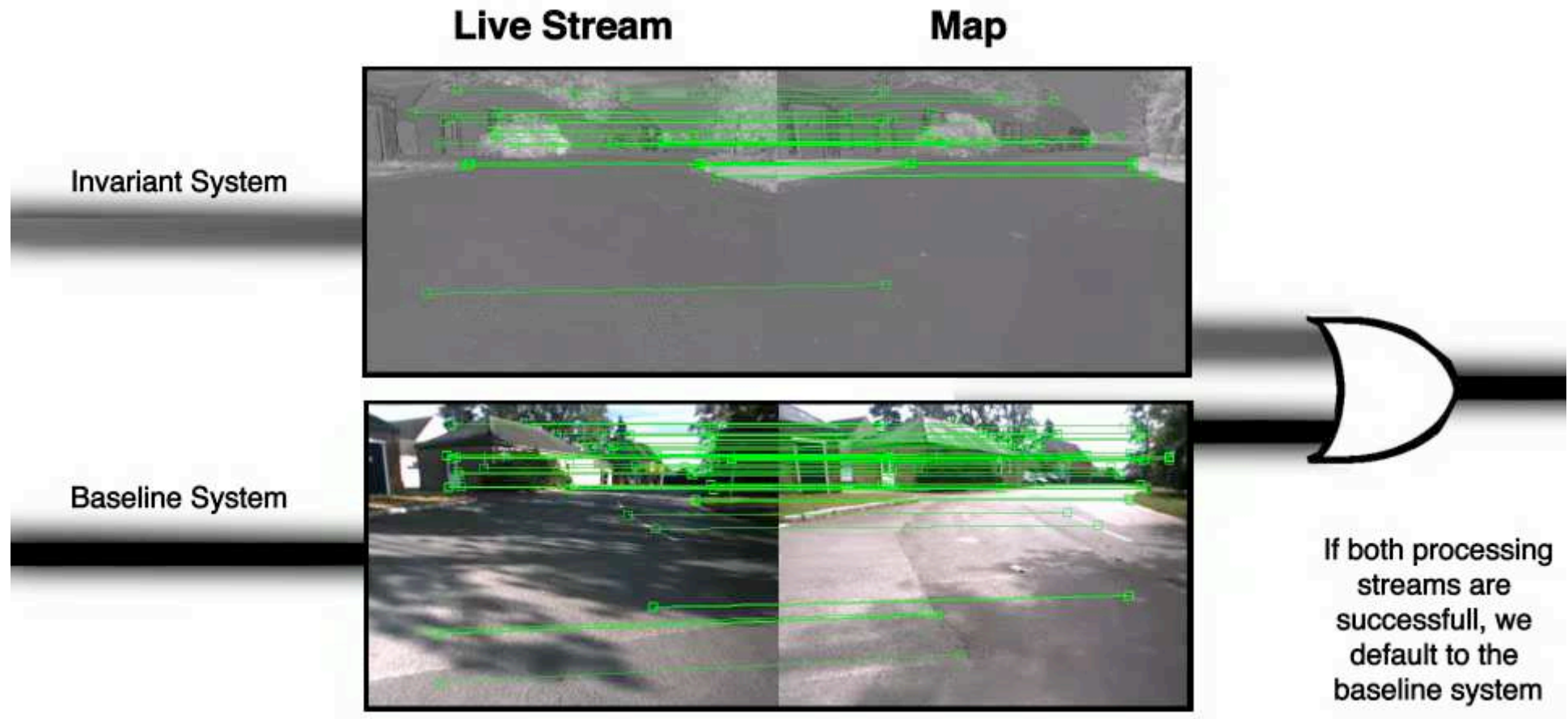
- Illumination invariant transforms fails:
 - At night (non-Black Body illumination)
 - When any channel is under/over exposed

Use Illumination Invariance as another Experience

Our Approach: Run an illumination-invariant image stream in parallel



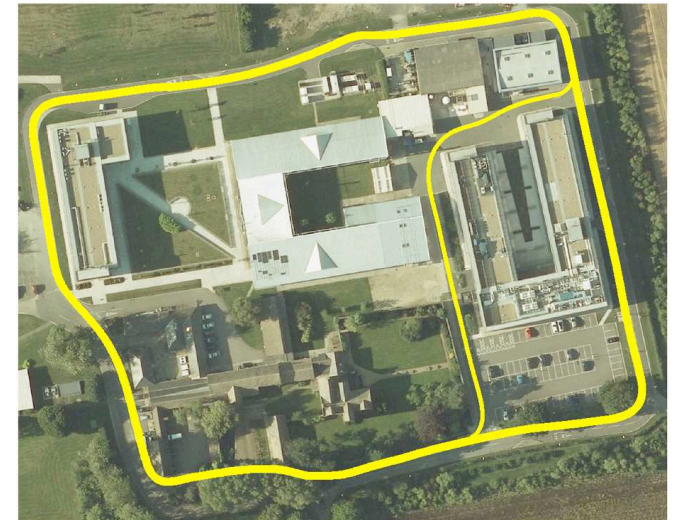
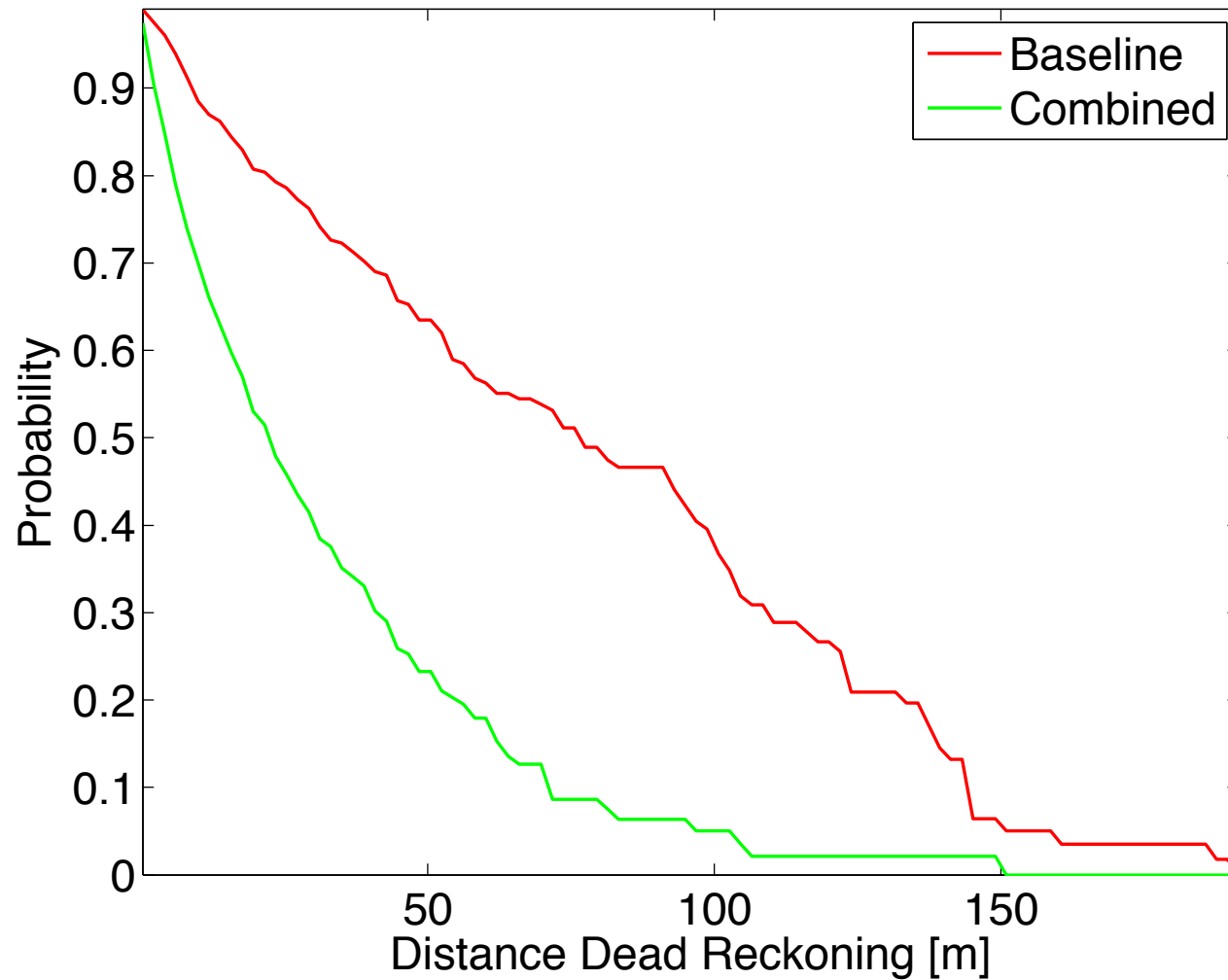
McManus, C., Churchill, W., Maddern, W., Stewart, A. D., & Newman, P., “*Shady Dealings: Robust, Long-Term Visual Localisation using Illumination Invariance*”, ICRA 2014



- Invariant system produces noisier estimates and a bias
- Invariant & baseline have *complimentary* failure modes
 - Combined system is more robust

Real Data - 3 Months, 10.5km, 15 loops

Probability of Distance Travelled using Dead Reckoning after Localisation Failure



Exploiting Prior Survey Data for Localisation



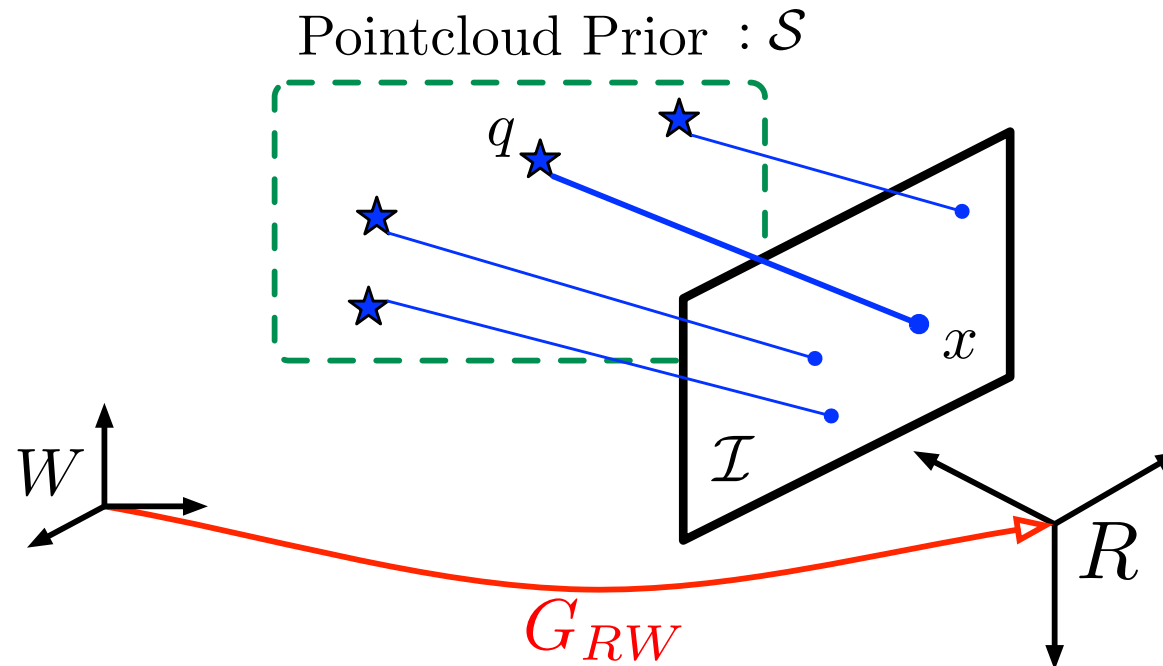
- How could we localise using monocular cameras if we had a 3D Scene Prior?
 - Build priors infrequently with expensive survey vehicles

Problem Formulation

Assumptions

- Operating environment has been completely mapped by a survey vehicle with a 3D lidar, cameras & INS
 - Pointcloud prior: $\mathcal{S}$ captures structure and reference appearance (i.e. exemplar colours for points)
 - Robot to be localised has monocular cameras
 - $\mathcal{S}$ has been preprocessed offline to improve its suitability for localisation (e.g. removing dynamic objects)
-
- **No** assumptions about external conditions experienced or sensor configurations between survey vehicle & robot

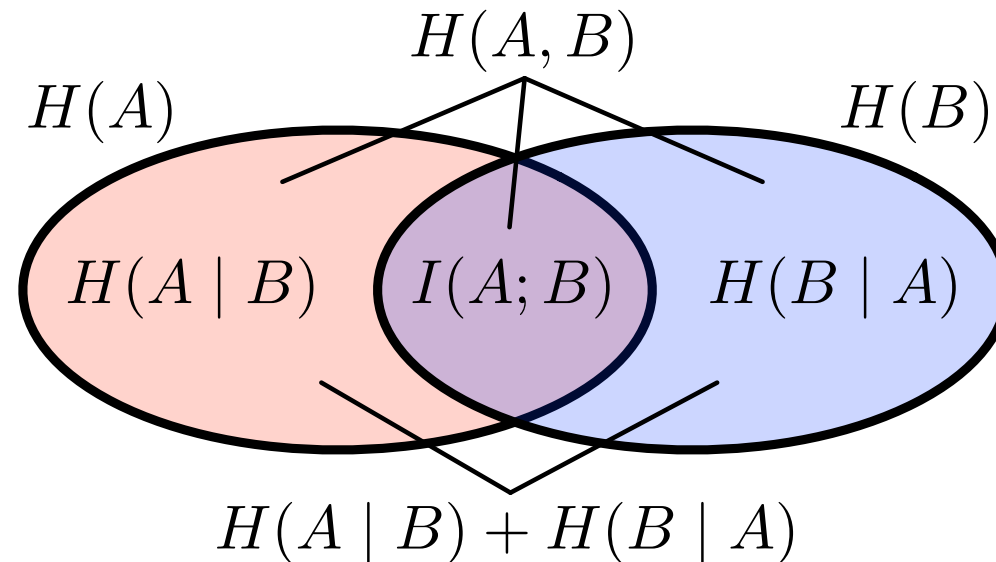
Formulation of Optimisation



Objective Function

$$\begin{aligned} \max_{G_{RW}} : & f_{\text{similarity}} \left(\underbrace{\mathcal{I}(x)}_{\text{Live Appearance}}, \underbrace{\mathcal{I}_S(x)}_{\text{Prior Appearance}} \right) \\ & \equiv \mathcal{I} \left(\mathcal{P}(q, \mathbf{G}_{RW}, \kappa) \right) \end{aligned}$$

What is a good choice for $f_{\text{similarity}}$?



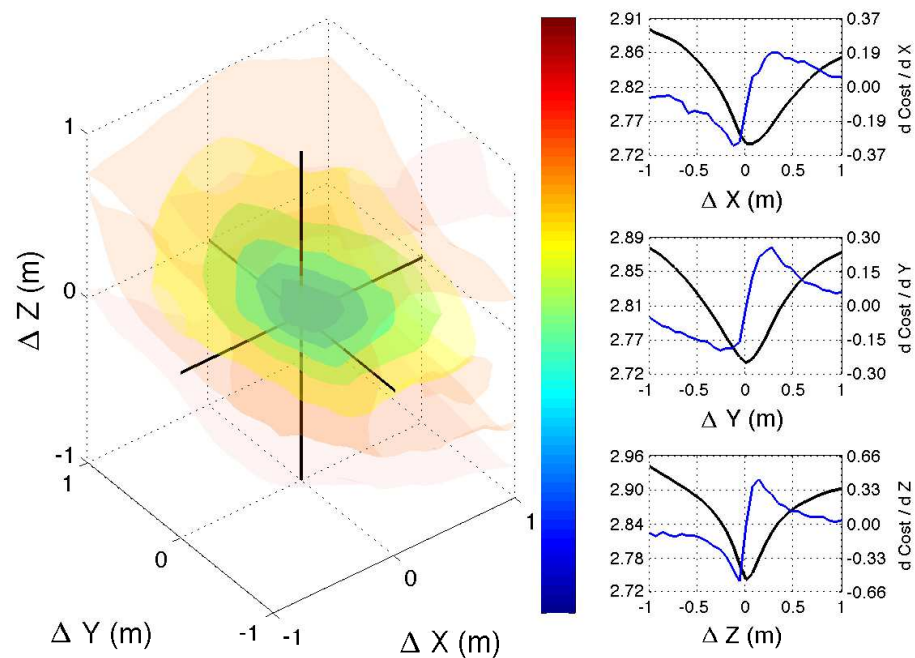
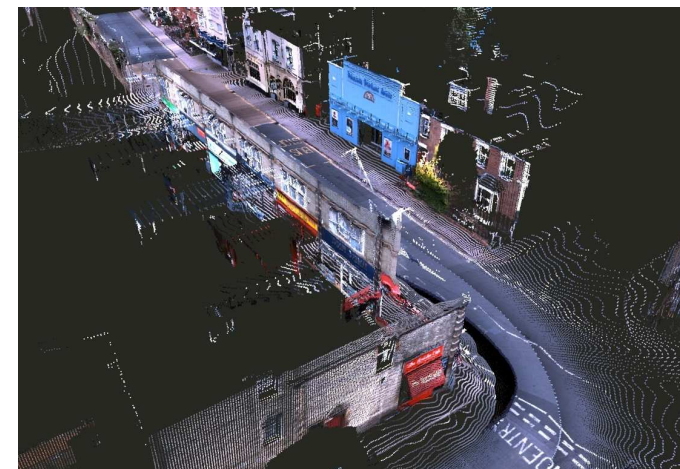
Normalised Information Distance (NID)*

$$NID(A, B) \equiv \frac{H(A | B) + H(B | A)}{H(A, B)}$$

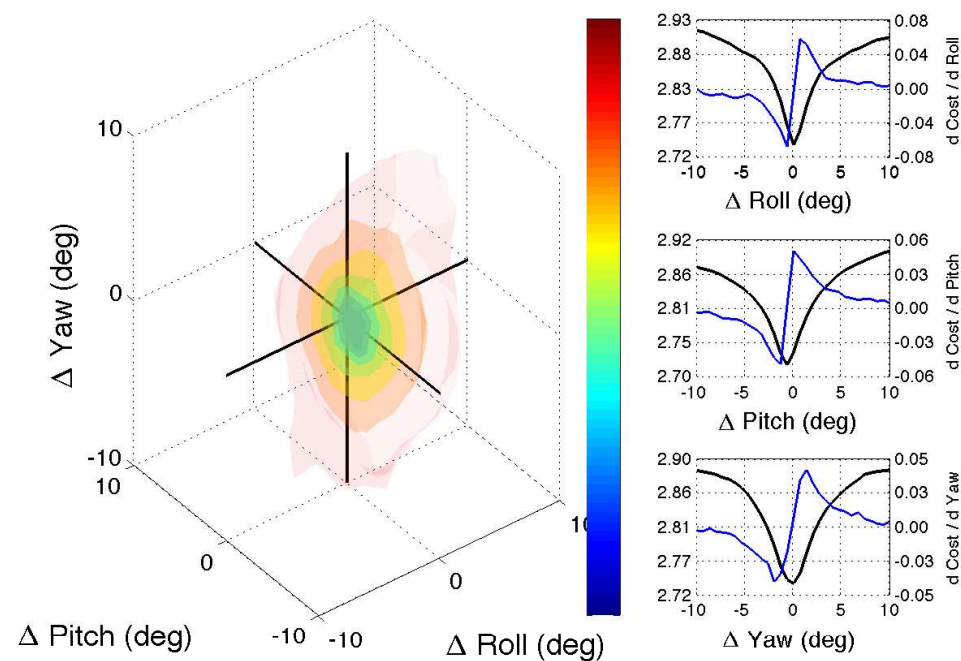
• NID is a **true metric** $\rightarrow$ triangle inequality

* M. Li, X. Chen, X. Li, B. Ma, and P. Vitanyi, "The similarity metric," *Information Theory, IEEE Transactions on*, vol. 50, pp. 3250 – 3264, dec. 2004

Localisation using the Appearance of Prior Structure

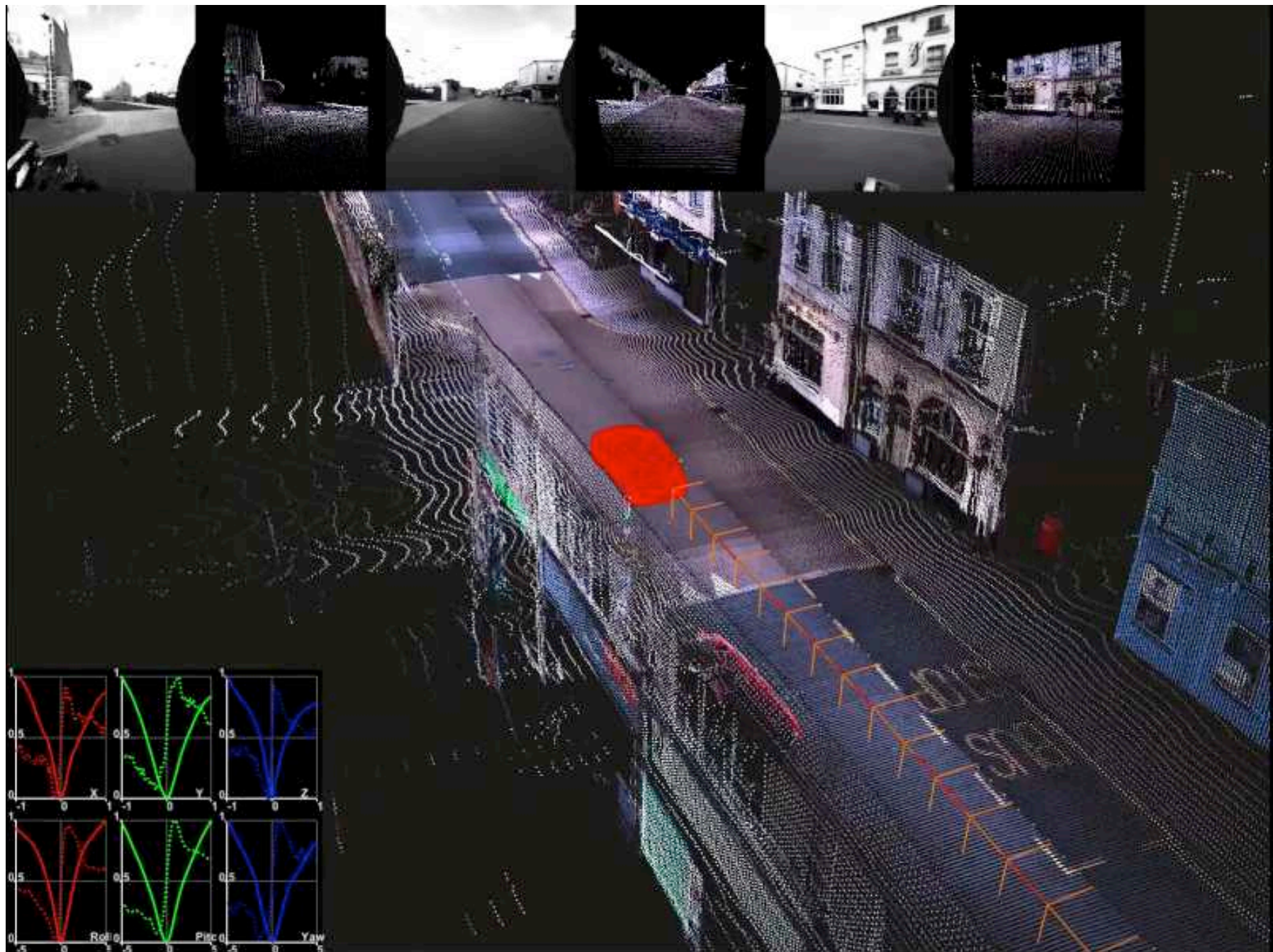


Δ Translation



Δ Orientation

Localisation using the Appearance of Prior Structure





<http://www.robots.ox.ac.uk/~mobile>