Indoor Localisation with Regression Networks and Place Cell Models

Jose Rivera-Rubio[1]
http://www.bicv.org
Ioannis Alexiou[2]
Anil A. Bharath[1]

[1] Biologically Inspired Computer Vision Group
Imperial College London
London, UK

[2] Computer Vision Research Group
Queen Mary University of London
London, UK

Abstract

Animals use a variety of environmental cues in order to recognise their location. One of the key behaviours found in a certain type of biological neuron – known as place cells – is a rate-coding effect: a neuron's rate of firing decreases with distance from some landmark location. In this work, we used visual information from wearable and hand-held cameras in order to reproduce this rate-coding effect in artificial place cells (APCs). The accuracy of localisation using these APCs was evaluated using different visual descriptors and different place cell widths. Simple localisation using APCs was feasible by noting the identity of the APC yielding the maximum response. We also propose using joint coding within a number of automatically defined APCs as a population code for self-localisation. Using both approaches we were able to demonstrate good self-localisation from very small images taken in indoor settings. The error performance using APCs is favourable when compared with ground-truth and LSD-SLAM, even without the use of a motion model.

1 Introduction

This work explores the potential of location-sensitive computational units to achieve self-localisation by using the images from wearable or hand-held cameras. Our approach is motivated by studies of biological vision which suggest that, in many species, multiple strategies are employed to help animals self-localise. Examples include the use of the visual horizon and path integration in ants [18] and the use of optic flow in insects and birds [4, 12]. Indeed, multiple computational approaches appear to be simultaneously at work even within a single species and for one sensing modality, such as vision. Such approaches can be successfully transferred from biology into computer vision; see for example, the combined use of optic flow and image descriptors that were suggested to mimic the visual homing system of insects [26].

Biological place cells display location-dependent firing. We describe how to mimic place cell behaviour from appearance information in order to provide localisation within an indoor environment. In terms of computer vision, our research question might be articulated as:

Pages 147.1-147.12
DOI: https://dx.doi.org/10.5244/C.29.147

Given a video sequence taken by a person as they walk, can we generate artificial place cell responses that are able to localise that person with respect to previous journeys?

1.1 Summary of Contributions

There are three primary contributions of this article. First, we suggest methods for appearance-based localisation by mimicking the behaviour of hippocampal place cells; the computational units of this behaviour – Artificial Place Cells (APCs) – are able to use distinctive information from image queries based on the recall of previously visited places. Secondly, we describe a complete pipeline of visual localisation that incorporates a decoder for the APC activity; this decoder takes the form of a Generalized Regression Neural Network (GRNN). Finally, we provide an evaluation of the proposed localisation method using a publicly available database of indoor sequences containing more than 120,000 frames and 3 km in length. Results suggest that sub-metre localisation error is achievable within this dataset.

2 Background

Navigation is one of the more complex tasks performed by animals; it involves integration of multiple sensory inputs, combining past information (memory) and also the execution of physical actions to perform navigation movements.

John O'Keefe, together with Edvard and May-Britt Moser, received the Nobel Prize in Physiology or Medicine (2014) for their discovery and subsequent elucidation of place and, subsequently, grid cells in the brain, respectively. Their early findings suggested that navigation in the moving rat is based on a cognitive map of the environment; in this map, a set of landmarks is created, and the spatial relationship between these is used to navigate [11]. The implications are interesting, partly for the potential importance in understanding some forms of dementia, but also because the cells encoding an organism's own spatial position are found in one of the less understood areas of mammalian brains: the hippocampus, with a key role in memory, and its adjacent brain areas.

2.1 Biological Place Cells (BPCs)

Place cells are special types of neurons located in the hippocampus which attain higher-than-average firing when an animal "recognises" a particular place in its environment. Grid cells, located next to the hippocampus in the entorhinal cortex, provide the brain with a reference system for navigation, a "grid" that is speculated to be used as a form of coordinate system for the creation of spatial maps. Although the existence of place and grid cells is without question, the computational description of what the cells do is debatable. However, the combination of place and grid cells within neural circuitry appears crucial for the execution of navigation tasks.

The physical region within a spatial environment over which a given place cell shows elevated firing is sometimes referred to as its *place field*, though the term may also refer to the mapping between an animal's location and a cell's firing rate. The combination of multiple place fields yields a spatial map, and multiple spatial maps formed by combinations of place cell activity patterns are thought to be stored in the hippocampus. It is the unique

combination of place cell firing patterns in a specific order during movement that gives rise to a unique spatial representation [20] of a journey.

The evidence is that many cues, and perhaps even self-motion itself, may be involved in forming the observed location-selective response of biological place cells. The visual information captured by the eyes should be seen as only one of the many sensory and internal cues that lead to the spatially selective nature of biological place cell responses [10]. Nevertheless, in many animals, and certainly in humans and primates, vision is a particularly strong environmental cue to an organism's awareness of its location [9].

2.2 Biologically Inspired Visual Localisation

Biologically-inspired methods of localisation from image data are emerging; for example, a few computational models for place cell behaviour already exist, though they are often rooted in dynamical systems [5],[3]. Some models of place cells use attractor properties of recurrent networks [24], [17]. Whilst interesting and valuable, the role of sensory input is marginalized in these models, a key differentiator of the approach we propose here.

The boundary vector cell (BVC) model [2] is a popular computational model that describes place cell response, allowing predictions to be made and experimentally tested [6]. Whilst of substantial interest in computational neuroscience, one criticism of the BVC model is – like the dynamical systems models – a lack of detail in explaining how sensory processing feeds into the computation. The closest work to the approach of this paper is Strosslin's [25], which uses a low-cost and computationally efficient model to navigate a robot. However, place cell behaviour – in terms of firing rates that vary with the position to some reference location – is not demonstrated except through the policy of a robot and its navigation attempts. The visual processing is also rather limited, using techniques that are also employed in SLAM algorithms [1] to track specific scene locations.

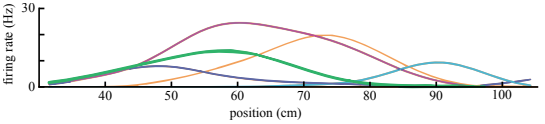
Milford and colleagues have also applied SLAM techniques to biologically inspired localisation systems, setting a seminal precedent with RatSLAM [16], a persistent navigation and mapping model based on modelling the hippocampus of rodents. RatSLAM continuously performs SLAM while simultaneously interacting with other navigation systems, such as odometry and landmark detection. More recent work [15] has been focused on taking RatSLAM to larger scales and incorporating more complex visual landmark detection models.

3 Artificial Place Cells (APCs)

3.1 Modelling a Single Place Cell: the Tuning Curve Encoder

Given a series of video frames extracted from footage recorded during indoor navigation, we made use of the visual path concept [14, 19, 22], to perform matching, or association, between locations of a physical environment being traversed and a database of previously captured journeys.

Our proposal – an evolution of that suggested for sensor and WiFi-based localisation – is that an appearance-based method using visual features could be easily mined to create a form of *virtual landmarks*. Such landmarks could be used to retrieve similar image locations around the locale of a landmark. The similarity scores obtained from appearance-based comparison methods, applied between sequences of frames of a journey and these virtual landmarks, should exhibit a behaviour that is similar to those recorded in mammalian place cells



(a) Firing rates of five BPC recordings covering different place fields of moving rats. Different colours represent different place cells. Adapted from [7].

Method	SIFT	DSIFT	SF-GABOR	ST-GABOR	ST-GAUSS
ST	No	No	No	Yes	Yes
Dense	No	Yes	Yes	Yes	Yes
Dim.	128	128	136	221	136

(b) Summary of the main properties of the different descriptors used. **ST**: Spatio-temporal, **SF**: Single-Frame

Figure 1: (Left) several BPC firing rate curves. (Right) Outline properties of the descriptors used in this study.

(Fig. 1(a)). In other words, one should obtain high scores when locations are visually similar or spatially close, and low ones when they are dissimilar, with a concave behaviour of scores with distance to the landmark. Requiring such behaviour of location sensing computational units would make them close to the behaviour of biological place cells.

In order to model the place cell behaviour based on the visual input provided by cameras, we first need a way of describing the collection of patches in an image near to a virtual landmark, and a means of comparing two images through their individual patch similarities.

We used visual features based on a set of publicly available keypoint and patch-based descriptor techniques: keypoint-SIFT (SIFT) [13], dense-SIFT (DSIFT) [27], single-frame (SF-GABOR) and spatio-temporal (ST-GABOR) Gabor based descriptors, and spatio-temporal Gaussian based descriptors [22]. Table 1(b) summarizes the techniques and shows the number of elements of each descriptor. For the comparison with a state-of-the-art SLAM method, we chose Engel's LSD-SLAM [8].

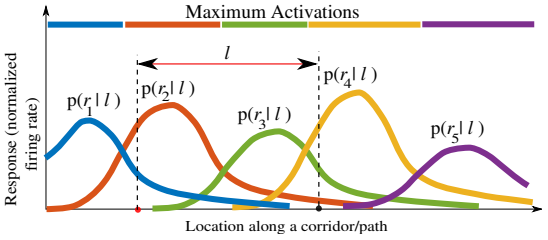
3.2 Encoding Frames of Video

Given a sequence of video frames, each of which is described by a collection of descriptor vectors, the task of comparing frames can be done by comparing vector pairs. Whilst a pair of frames captured in two different journeys might differ for a number of reasons that include partial occlusion, motion blur, lighting changes and object/scene changes, one would expect that there might be some patches that are similar. However, in a 1 minute video, one might capture 1,500 frames, with each frame containing several thousand patch descriptors, each of them a high-dimensional vector.

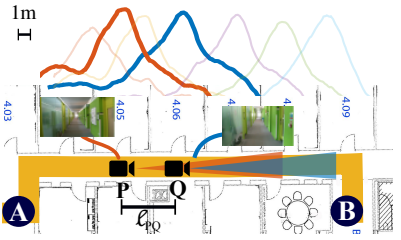
One solution to this "curse of dimensionality" is to use vector quantisation (VQ, 4^{th} module of the pipeline described in Fig. 5), creating a mapping (codebook) that reduces each of the descriptors to a single number representing the identifier of an area of descriptor space. We applied the k-means algorithm with $k = 400$ to samples of video data acquired from corridors in order to generate the VQ codebook. Patch descriptors were L_2 normalized prior to applying the k-means algorithm, then each frame was separately encoded into a 400-element Frame-Encoding Vector (FEV).

3.3 Comparing Frames

In order to model place cell behaviour, we need to map pairs of frames onto a scalar value that is analogous – perhaps after a non-linearity and affine scaling – to a firing rate. Consider



(a) Each curve, known as a place cell tuning curve, represents the firing rate of an individual place cell to a position along a one-dimensional path. The simplest decoding of place cell responses is to find the cell which displays maximum response, localizing a camera as being within the "receptive field" of that place cell (coloured horizontal bars). A more sophisticated decoding method employs a regression network.



(b) Illustration of prototype place cell behaviour produced by comparing FEVs from a series of frames to those at several landmark locations with an appropriate metric. The curves shown here do not include thresholding (see text).

Figure 2: Hypothetical (desired) responses of APCs (left), and diagram of place cell generation from image comparisons (right).

two image frames that are captured at positions P and Q, and are spaced a distance, ℓ_{PQ}, apart (Fig. 2(b)). As the distance ℓ_{PQ} varies, the mapping should yield a smoothly varying result. In addition, biological place cell (BPC) behaviour often displays a concave relationship between firing rate and distance from the location of peak response (often, the location of some landmark) as illustrated in Fig. 2(a) and Fig. 2(b).

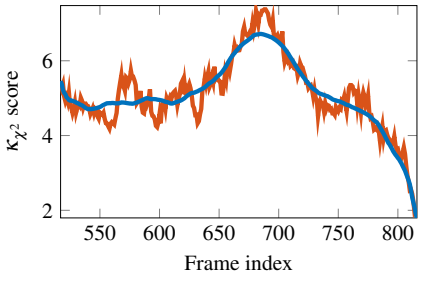
One way to mimic the behaviour of BPCs within APCs is to introduce a *kernel* function that maps a pair of FEVs onto a positive scalar value. We used the following mapping between two vectors $\mathbf{v}_a$ and $\mathbf{v}_b$:

$$\kappa_{\chi^2}(\mathbf{v}_a, \mathbf{v}_b) = \sum_{j=1}^{400} \frac{v_a(j) \cdot v_b(j)}{v_a(j) + v_b(j)} \tag{1}$$

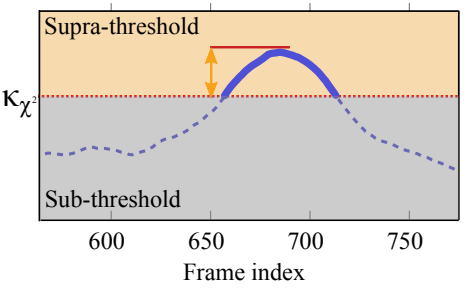
This maps the two FEVs onto a scalar value that takes a maximum when the two vectors are identical. The behaviour of this kernel mapping between a fixed frame and a series of frames from a sequence is shown in Fig. 3(a). As one moves along the horizontal axis from left to right, vectors from each frame in a section of a video sequence are compared against those of a *reference* frame (virtual visual landmark) from another sequence. The location of the reference frame corresponds to the location in the new sequence at which one observes the peak in the curve (around frame 690) of Fig. 3(a); this location corresponds to the effective position of the virtual landmark, which is the field of view in front of the camera at that (frame) location. The variability of responses is high, but it is not unlike the types of variation found in biological place cell behaviour.

4 Creating Place Fields

The similarity function κ_{χ^2} is an important component in producing APC behaviour from FEVs. However, regions of the APC response curve which are almost flat contain very little information from the point of inferring the position of a stimulus that elicits some response. In other words, regions of the curve where the gradient with respect to distance to the peak of



(a) Single APC tuning curve (raw measurements in red) and a smoothed version (blue trace). The reference frame (virtual visual landmark) is located around frame 690. The APC response must be thresholded before using it for accurate position inference.



(b) Sub-threshold and supra-threshold regions can be identified by setting a threshold on the amplitude of the κ_{χ^2} similarity measure; the height of the threshold controls the support region of supra-threshold region of an artificial place cell.

Figure 3: (Left) A single APC directly obtained from the κ_{χ^2} similarity metric. (Right) How hard-thresholding may be used to obtain a final APC tuning curve.

the curve, ℓ, is small, convey relatively little information about the camera's location. Other regions of the curve show rapid change with distance from the peak. In order to synthesize useful "tuning curves" for APCs, we need to define sub- and supra-threshold regions for each APC that we wish to define (Fig. 3(b)).

The place cell is modelled by extracting the Frame-Encoding Vector (FEV), $\mathbf{v}_{r_i}$, when the camera is at position ℓ_i from one or more reference journeys and calculating the supra-threshold response to some frame $\mathbf{v}_\ell$ acquired at location ℓ. As ℓ is varied, $\kappa_{\chi^2}(\mathbf{v}_\ell, \mathbf{v}_{r_i})$ changes accordingly. The set of supra-thresholded response curves, $r_i(\ell)$, is generated using:

$$r_i(\ell) = U(\kappa_{\chi^2}(\mathbf{v}_\ell, \mathbf{v}_{r_i}) - T_i) \cdot \kappa_{\chi^2}(\mathbf{v}_\ell, \mathbf{v}_{r_i}) \tag{2}$$

where $U(\cdot)$ represents the unit step function and T_i represents a suitable threshold. Curves acquired by averaging responses from several journeys with respect to the same APC location may be referred to as an APC *tuning curve*.

By defining APCs at regular intervals along a corridor, a simple population code for location can be devised. In Fig. 4, the *average* response from each of several such APC cell responses is plotted along the length of one corridor. These curves are produced by setting APCs to be spaced every 4 m within the corridor, and constructing the average APC responses. In Fig. 4, ground truth was used to register the curves for illustrative purposes.

The responses of the collection of APCs provide important information as a population code [21]. By changing the threshold level, the support region of the APCs can be altered, leading to greater or lesser degrees of overlap, and (see Section 5), different performance.

4.1 Location from APC Activations

Conceptually, given a series of APC responses to visual cues of a person's location along some journey – illustrated in Fig. 2(a) – there are two obvious ways of estimating location, ℓ. The first is simply to use the APC which displays maximum activation (firing rate) as a rough indicator of where the person is. That is, given a set of activations $p(r_i|\ell)$, the location of the camera that captured a particular image frame is provided by the index, i associated

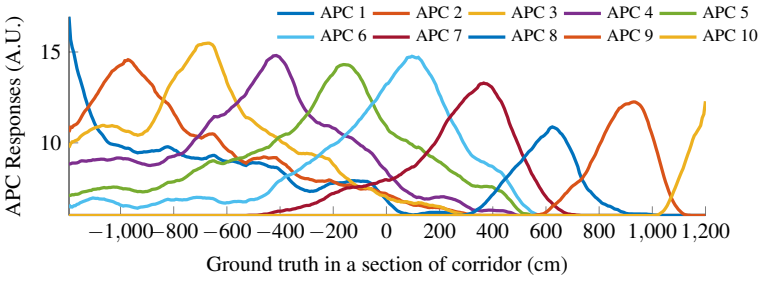


Figure 4: Responses from APCs in the RSM dataset (see Section 5). The APCs are defined every 4 m within a corridor, using ground truth information. Once defined, these APCs provide two different methods of localisation.

with maximizing $p(r_i|\ell)$. This provides a precision that is limited by the width the APCs, but requires little more than ensuring that place cell responses are reasonably well-separated.

The second technique to infer location achieves more accurate localisation of a camera from its captured visual data by using the joint distribution $p(\mathbf{r}|\ell)$, of APC responses, $\mathbf{r}$ to infer location ℓ relative to some designated ground truth. We use a single index, i, to refer to the response of a unique place cell, r_i.

First, a rough location may be identified by using $\texttt{arg max}(r_i)$ over the index, i; then, the responses of neighbouring APCs can be used to obtain sub-cell localisation. In order to apply this principle, one needs sufficiently accurate estimates of $p(\mathbf{r}|\ell) = p(r_1, r_2, r_3, ..., r_C|\ell)$, where r_C is the total number of place cells in some region of a path. Given several active cells that are a subset of all place cells in a location, sub-APC localisation is possible using APC responses from previous journeys using empirical Bayes' techniques. For example, if three cells are active, the chain rule can be used to obtain successively refined estimates of ℓ:

$$\begin{aligned} p(\ell|\mathbf{r}) &\propto p(r_3, r_4, r_5|\ell)p(\ell) \\ &\propto p(r_3|r_4, r_5, \ell) \times p(r_4|r_5, \ell) \times p(r_5|\ell) \times p(\ell) \end{aligned} \tag{3}$$

so that the responses of spatially close APCs can be used to infer sub-APC position. If the width of an APC is set to around 2 m, localisation of the order of tens of centimetres is plausible.

4.2 Overview of the System

In the experimental work to be discussed in the next section, a Generalized Regression Neural Network was used to provide sub-APC position estimates, obviating the need to construct *ad-hoc* empirical estimators. This regression network consists of two-layers, and uses radial-basis functions. The responses from $C = 16$ place cells were input to the network, and ground truth of location within a section of corridor – up to 4 m long – used to train it as a regressor. In all experiments, dictionary generation was performed independently of the APC responses used in training the regression network.

An overview of the pipeline for processing the frames to generate APCs is shown in Fig. 5. The performance of different methods will be discussed in Section 5.

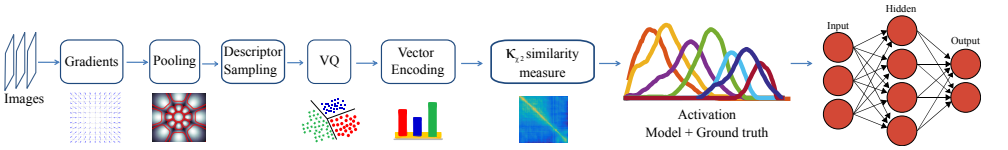


Figure 5: Overview of the training pipeline. The diagram of the neural network is merely illustrative, and does not represent the real GRNN architecture used.

5 Experiments and Results

In order to test the APC concept, we performed a series of experiments using the publicly available RSM dataset [23]. This dataset contains visual data of more than 3 km of indoor journeys acquired with two devices: a hand-held Nexus 4 and a wearable Google Glass. Corridors varied in length between 32 m and 60 m and the sequences comprise more than 120,000 frames with ground truth acquired with surveying equipment. The original resolution of the dataset is 1280×720 pixels per frame which we down-sampled to 208×117.

5.1 APC-level Localisation

The first method used to infer location – based on identifying the APC with maximum activity – was tested in several corridors of the experimental data. First, a visualization showing the locations of APCs with maximum activation is shown in Fig. 6(a); these are indicated on the floor plan of one of the building sections that was used to conduct the experiment. Locations are staggered across the width of the corridor in order to visualize the individual activations of the 8 APCs defined within this corridor. The second technique to estimate location relies on the overlap of the responses from APCs and the use of a GRNN. Table 1 compares the two localisation techniques for different methods of descriptor generation. A dictionary based on a single device type was used, but all the combinations of remaining passes were submitted as queries. The neural network regressor shows better results, achieving errors as low as 2.49 m for SF-GABOR, even with the majority of the queries coming from a different device that was not used to learn the dictionary. Very low errors were observed using a single device, as may be seen in Fig. 6(b). The performance of LSD-SLAM, when tested on the same sequences, is also reported, but tracking was lost in roughly 40% of sequences. This is probably because LSD-SLAM performs best on sequences acquired with global-shutter, fish-eye lenses. The RSM dataset does not use such cameras. A tracking recovery exception catch was therefore implemented to keep LSD-SLAM running.

5.2 Sub-APC Localisation

By arranging for APC responses to be up to 4 m in width, the responses from several cells can be used to perform accurate inference of spatial position using a single section of corridor. The success of this technique is illustrated in Fig. 6(b). Note that average absolute errors are very small compared to distances traversed.

	Sub-APC Localisation				APC-level Localisation			
Method	$\mu_{\|\varepsilon\|}$ (m)	$\sigma_{\|\varepsilon\|}$ (m)	Min (m)	Max (m)	$\mu_{\|\varepsilon\|}$ (m)	$\sigma_{\|\varepsilon\|}$ (m)	Min (m)	Max (m)
SIFT	3.42	2.72	0.38	6.47	4.48	5.14	0. 34	7.56
DSIFT	2.78	2.45	0.47	6.54	4.10	5.64	0.31	6.10
SF-GABOR	2.49	2.07	0.46	5.60	4.81	7.11	0.68	8.47
ST-GABOR	7.94	5.54	2.63	10.45	9.65	9.19	3.67	13.13
ST-GAUSS	3.18	2.64	0.45	7.17	3.79	5.44	0.69	8.12
LSD-SLAM$^{(*)}$	$\mu_{\|\varepsilon\|}$ = 2.48 m		$\sigma_{\|\varepsilon\|}$ = 2.37 m		Min: 1.21 m	Max: 3.20 m		

Table 1: Absolute error evaluation when using a larger number (40) of APCs of small spatial support (0.61 m), using `arg max()` to infer spatial position, in contrast to using fewer (16) but larger APCs with substantial overlap and the regression network (sub-APC). The comparison with a state of the art SLAM method (LSD-SLAM) is also included.$^{(*)}$ LSD-SLAM performance is positively affected by the tracking recovery exception, which reduces the error drift by reseting the odometry calculation and the error of the pose-graph optimisation.

5.3 Parameter Tuning

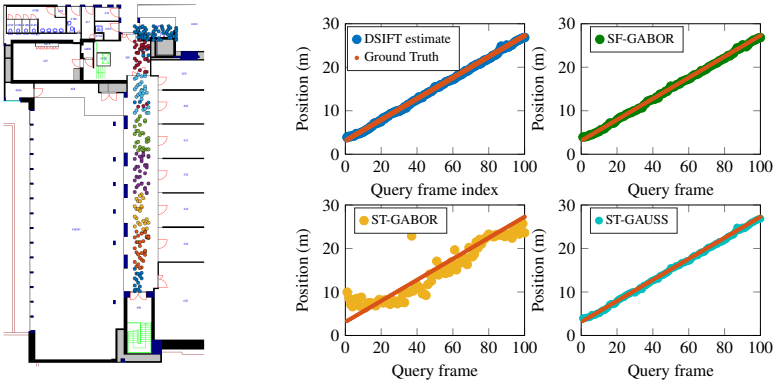
The GRNN relied on the overlap of the responses of the place cells to perform location inference. By varying the threshold beyond which place cells are considered active, it is possible to have very long-tailed APC responses, spanning several metres. This yields APC behaviour that is similar to that of rate-coding in place cells observed in biology. The blue trace on Fig. 7(a) shows an example of the average cell width in metres as the threshold on the κ_{χ^2} comparison metric is varied. The red trace illustrates the average error, also in metres, that is obtained from the GRNN regressor. Having low APC response overlap deprives the network of sufficient non-zero inputs, leading to poor accuracy in position inference.

One of the key problems with using different cameras – even if they are calibrated – is the difference in field of view of one imaging device with respect to another. A Nexus 4 smartphone and a wearable Google Glass were used to conduct experiments into the effect of the dictionary on the localisation using the sub-APC (GRNN) approach.

A set of 10 journeys through one of the corridors of the RSM building with different devices was taken for this experiment. Some of the journeys were included in the database, others were used to conduct queries. The partitioning of the data was permuted, varying the number of journeys in the database and the number of query journeys kept out of the database. Journeys which were used for queries did not contribute to the learning of the dictionary, which was repeated for each permutation. Fig. 7(b) shows the difference in absolute error of Nexus and Glass queries when only passes from one device (in this case, the Nexus) were used for the dictionary learning.

6 Conclusion

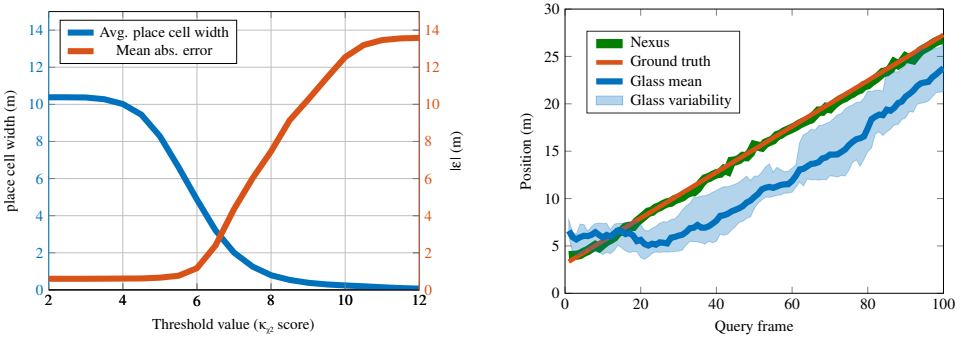
To our knowledge, there has been no reported computational architecture that reproduces biological place cell responses in the form of tuning curves from video sequences, and there have been no other demonstrable computational models for place cell positional encoding that can be applied to the recorded video sequences of the form we used in this work. Furthermore, competing camera-based localisation techniques such as SLAM rely on relative object camera motion to infer structure and to then either perform odometry or to estimate position: motion is therefore a requirement. The place cell model that we have proposed and evaluated in this paper does not require motion at location estimation time: a single frame



(a) Using ground truth from a surveyor's wheel, activations from APCs are overlaid onto the floor plan in which video data was acquired. Different colours refer to individual APCs.

(b) Sub-APC location estimate comparison. Using broad APC tuning curves, very accurate localisation can be achieved within a section of corridor. For this corridor and for this journey, absolute localisation errors range from below one metre to 1.49 m when only sequences from the Nexus 4 were used. Ground truth is shown in red.

Figure 6: (Left) a visualization of APC-level localisation over a floor plan. (Right) Sub-APC localisation results with different patch descriptor methods.



(a) Effect of varying the threshold on place cell width, and average absolute error in metres.

(b) Multi-device sub-APC localisation test. Note the substantially worse cross-device performance.

Figure 7: Evaluation of sub-APC localisation when using different thresholds and query devices.

yields a hypothetical location.

In conclusion, this work demonstrates that computational models of place cells can provide effective estimates of camera location without relying on tracking or construction of a geometric model of the local environment. Such techniques, although simple, can match and certainly complement the more sophisticated position inference approaches used in computer vision.

Acknowledgements Jose Rivera-Rubio's research is supported by an EPSRC DTP scholarship and The National Archives through the "Perceptually Similar" project. The experiments and simulations were run using the Imperial College Neurotechnology CDT servers.

References

[1] Pablo Fernández Alcantarilla, Luis Miguel Bergasa, and Frank Dellaert. Visual odometry priors for robust ekf-slam. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3501–3506. IEEE, 2010.

[2] Caswell Barry, Colin Lever, Robin Hayman, Tom Hartley, Stephen Burton, John O'Keefe, Kate Jeffery, and N Burgess. The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, 17(1-2):71–98, 2006.

[3] William Bechtel. Investigating neural representations: the tale of place cells. *Synthese*, pages 1–35, 2013.

[4] Partha S Bhagavatula, Charles Claudianos, Michael R Ibbotson, and Mandyam V Srinivasan. Optic flow cues guide flight in birds. *Current Biology*, 21(21):1794–1799, 2011.

[5] Hugh T Blair, Kishan Gupta, and Kechen Zhang. Conversion of a phase-to a rate-coded position signal by a three-stage model of theta cells, grid cells, and place cells. *Hippocampus*, 18(12):1239–1255, 2008.

[6] Neil Burgess, Andrew Jackson, Tom Hartley, and John O'Keefe. Predictions derived from modelling the hippocampal role in navigation. *Biological cybernetics*, 83(3):301–312, 2000.

[7] George Dragoi and Susumu Tonegawa. Selection of preconfigured cell assemblies for representation of novel spatial experiences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120522, 2014.

[8] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *Computer Vision–ECCV 2014*, pages 834–849. Springer, 2014.

[9] Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

[10] Demis Hassabis, Carlton Chu, Geraint Rees, Nikolaus Weiskopf, Peter D Molyneux, and Eleanor A Maguire. Decoding neuronal ensembles in the human hippocampus. *Current Biology*, 19(7):546–554, 2009.

[11] John O Keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Clarendon Press Oxford, 1978.

[12] Holger G Krapp. Neuronal matched filters for optic flow processing in flying insects. *International review of neurobiology*, 44:93–120, 2000.

[13] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[14] Yoshio Matsumoto, Masayuki Inaba, and Hirochika Inoue. Visual navigation using view-sequenced route representation. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, volume 1, pages 83–88. IEEE, 1996.

[15] Michael J Milford and Gordon Fraser Wyeth. Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1643–1649. IEEE, 2012.

[16] Michael J Milford, Gordon F Wyeth, and David Prasser. Ratslam: a hippocampal model for simultaneous localization and mapping. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 403–408. IEEE, 2004.

[17] Edvard I Moser, Emilio Kropff, and May-Britt Moser. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, 31:69–89, 2008.

[18] Ajay Narendra, Sarah Gourmaud, and Jochen Zeil. Mapping the navigational knowledge of individually foraging ants, myrmecia croslandi. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765):20130683, 2013.

[19] Takayuki Ohno, Akihisa Ohya, and S Yuta. Autonomous navigation for mobile robots referring pre-recorded image sequence. In *Intelligent Robots and Systems' 96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on*, volume 2, pages 672–679. IEEE, 1996.

[20] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175, 1971.

[21] Alexandre Pouget, Peter Dayan, and Richard Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000.

[22] Jose Rivera-Rubio, Ioannis Alexiou, Luke Dickens, Riccardo Secoli, Emil Lupu, and Anil A Bharath. Associating locations from wearable cameras. In *Proceedings of the British Machine Vision Conference*, 2014.

[23] Jose Rivera-Rubio, Ioannis Alexiou, and Anil A Bharath. Appearance-based indoor localization: A comparison of patch descriptor performance. *Pattern Recognition Letters*, 2015.

[24] SM Stringer, ET Rolls, TP Trappenberg, and IET De Araujo. Self-organizing continuous attractor networks and path integration: two-dimensional models of place cells. *Network: Computation in Neural Systems*, 13(4):429–446, 2002.

[25] Thomas Strösslin, Denis Sheynikhovich, Ricardo Chavarriaga, and Wulfram Gerstner. Robust self-localisation and navigation based on hippocampal place cells. *Neural networks*, 18(9):1125–1140, 2005.

[26] Andrew Vardy and Ralf Moller. Biologically plausible visual homing methods based on optical flow techniques. *Connection Science*, 17(1-2):47–89, 2005.

[27] Andrea Vedaldi and Brian Fulkerson. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia*, pages 1469–1472. ACM, 2010.