Unsupervised Online Learning of Condition-Invariant Images for Place Recognition

Stephanie Lowry[1], Gordon Wyeth[1,2] and Michael Milford[1,2]
[1]School of Electrical Engineering and Computer Science, Queensland University of Technology
[2]Australian Centre for Robotic Vision, Queensland University of Technology
stephanie.lowry@student.qut.edu.au

Abstract

This paper presents an online, unsupervised training algorithm enabling vision-based place recognition across a wide range of changing environmental conditions such as those caused by weather, seasons, and day-night cycles. The technique applies principal component analysis to distinguish between aspects of a location's appearance that are *condition-dependent* and those that are *condition-invariant*. Removing the dimensions associated with environmental conditions produces condition-invariant images that can be used by appearance-based place recognition methods. This approach has a unique benefit – it requires training images from only one type of environmental condition, unlike existing data-driven methods that require training images with labelled frame correspondences from two or more environmental conditions. The method is applied to two benchmark variable condition datasets. Performance is equivalent or superior to the current state of the art despite the lesser training requirements, and is demonstrated to generalise to previously unseen locations.

1 Introduction

Place recognition across changing conditions is a continuing challenge for vision-based robotic navigation systems. Vision-based systems struggle to identify the same place under different conditions, such as changes in lighting, seasons, or weather conditions [Valgren and Lilienthal, 2010]. Existing approaches use various processing techniques to improve performance but struggle to generalise to different locations or across different environmental conditions. Learning-based methods [Johns and Yang, 2013; Neubert *et al.*, 2013] are one of the most promising approaches but require appropriately curated training data spanning all possible environmental conditions and labelled frame correspondences across these conditions.

This paper proposes a novel concept in place recognition: to distinguish between *condition-dependent* aspects of an environment such as lighting and weather conditions, which are irrelevant for – and even detrimental to – place recognition, and *condition-invariant* conditions which make a place distinctive from other nearby places. We utilise Principal Component Analysis (PCA) [Hotelling, 1933; Pearson, 1901] to identify and remove condition-dependent aspects and produce condition-invariant images for use in place recognition.

Figure 1 demonstrates this concept using two images from the same location – one captured in winter and one captured in spring. The images generated using the first 100 PCA dimensions (second column) reflect the common aspects of each environmental condition (snow cover in winter and darker ground in spring). However, the images generated using only the later PCA dimensions (third column) appear more similar than the raw spring and winter images. In this paper we test the hypothesis that using the later PCA dimensions retains the distinctive elements of a location necessary for place recognition, while eliminating unhelpful condition-dependent information.

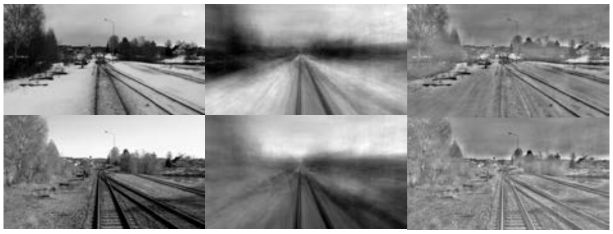


Figure 1. Grayscale images from the same location in winter (top row) and spring (bottom row). When a PCA decomposition is learned on a dataset, images formed using the first 100 dimensions (second column) tend to reflect the common aspects of the dataset (white snow in winter, darker ground in spring) but the images formed from later PCA dimensions (third column) contain more condition-invariant information. The first 100 PCA dimensions represent 88% of the variance in the training data.

The proposed system has two significant advantages over existing techniques. Firstly, the required information is learned online from environmental data, so the type of change occurring does not need to be known beforehand. The second advantage is that the training data can be generated online. The PCA decomposition represents a single environmental configuration, so the training data can be generated by sampling images from locations that are close together in time and place.

The paper proceeds as follows: Section 2 briefly summarises prior work in place recognition in changing environments, and discusses the application of PCA to learning from image data. The mathematical framework used is introduced in Section 3. Sections 4 and 5 experimentally test this approach across benchmark seasonal change and day-night datasets. The paper concludes with a discussion of the results in Section 6.

2 Prior work

The problem of environmental appearance change over time is significant for robot systems that use vision to localise, particularly those operating in outdoor environments where appearance change is often drastic. Section 2.1 discusses existing approaches to performing visual place recognition in changing environments. Section 2.2 looks at the role of principal component analysis in image processing, motivating its application here.

2.1 Visual place recognition in changing environments

Because of the immediacy of the lighting variance problem, substantial research has been dedicated to ameliorating its effects. Techniques that improve visual localisation in changing lighting conditions include methods of illumination invariant processing [Corke *et al.*, 2013; Maddern *et al.*, 2014]. Image matching techniques that use features such as SURF [Bay *et al.*, 2008] can be combined with geometric constraints to improve matching [Cadena and Neira, 2011; Cummins and Newman, 2011; Valgren and Lilienthal, 2010], but degrade rapidly as conditions change. A hardware-based solution to lighting-invariant place recognition is to use lidar (scanning laser-rangefinders) [McManus *et al.*, 2013] to create *camera-like* images that are not affected by lighting.

Many techniques for learning about change assume the system has access to multiple traverses of an environment at different times, and that the images from each traverse are perfectly aligned [Carlevaris-Bianco and Eustice, 2014; Johns and Yang, 2013; Lowry *et al.*, 2014; McManus *et al.*, 2014; Neubert *et al.*, 2013; Ranganathan *et al.*, 2013]. Such specialised training data cannot be generated by the robot itself but needs to be provided from an external source. Furthermore, none of these techniques have been shown to generalise easily to new locations.

Sequences of images can be used to match locations despite changes in lighting and weather conditions, or poor visibility [Badino *et al.*, 2012; Milford and Wyeth, 2012]. Sequential information can also be modelled as a network flow [Naseer *et al.*, 2014] that finds the least cost path through the observation data.

Visual place recognition systems based on single image matching rather than image sequence matching can use a two-stage process to recognise locations in changing conditions [Milford *et al.*, 2014]. The first stage compares at the image-level to select hypothesis images. The second stage uses higher-resolution images on which patch-matching and shift coherency testing is performed.

2.2 PCA for visual recognition tasks

PCA [Hotelling, 1933; Pearson, 1901] has been used for visual recognition tasks including face recognition [Sirovich and Kirby, 1987; Turk and Pentland, 1991] and place recognition in unchanging conditions [Artac *et al.*, 2002; Carreira *et al.*, 2014; Dudek and Jugessur, 2000; Liu and Zhang, 2012]. In each case the early PCA dimensions were kept and in many cases the later PCA dimensions were discarded. The technique presented in this paper differs from these approaches by discarding the early PCA dimensions and keeping the later ones. This approach is less common but an example can be found in the analysis of video surveillance footage from static cameras [Candès *et al.*, 2011; Chandrasekaran *et al.*, 2011]. In such cases, the background of the extracted images will stay very similar over time, and is likely to be identified by the early dimensions of a PCA decomposition. However, the interesting data of such footage is the motion of people in the foreground, which only appears occasionally and is likely to reside in the later dimensions.

The visual place recognition system presented here is conceptually similar to the background extraction problem – we want to extract the common but uninteresting condition-dependent information and use the rare and interesting location-specific details. The next section describes the application of PCA to this problem.

3 Approach

The approach proposed in this paper uses PCA to estimate the condition-dependent and condition-invariant aspects of a scene. PCA is a basis transformation – it transforms data "to a new set of variables…which are ordered so that the first *few* retain most of the variation present in *all* of the original variables" [Jolliffe, 2002, p.1]. In other words, the PCA transformation identifies the dimensions that contain the most common information within the training set, and the basis is ordered so that these dimensions come first.

This section summarises the mathematical framework required for calculating the principal components (Section 3.1) and describes how condition-invariant images can be generated (Section 3.2).

3.1 Determining the principal components

The principal component transformation of a set X is defined as the orthogonal transformation where the most variance is compressed into the least number of dimensions. It can be shown that this basis set is equal to the eigenvectors of the data set's covariance matrix X^TX [Jolliffe, 2002]. One way to calculate the principal components of a set is to use the singular value decomposition (SVD). The SVD of X is defined as:

$$X = USC^T \tag{1}$$

Here C is the set of eigenvectors of the dataset X^TX (X is real as it represents image data), and thus represents the principal components for X. Due to the nature of SVD, the matrix S is a diagonal matrix:

$$S = \begin{bmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_N \end{bmatrix} \tag{2}$$

Furthermore, the squares of the diagonal values of S, $e_1 = S_1^2$, $e_2 = S_2^2$, … are the eigenvalues corresponding to the eigenvectors C. Importantly, the eigenvalues are proportional to the amount of variance contained in the related dimension; that is, e_1 relates to the amount of variance in the first dimension C_1, e_2 relates to the amount of variance in the second dimension C_2, and so on. These eigenvalues are a measure of how much information is contained in each dimension, and are useful for understanding the underlying nature of the data.

3.2 Condition-invariant images using PCA

The PCA transformation described above can be applied to a training set of images I_{train}. The principal components C are calculated from I_{train} via the SVD calculation described above, that is:

$$I_{\mathrm{train}} = USC^T \tag{3}$$

Once calculated, C can be used to transform any image I into the principal component space, not just images from I_{train}. The projection of I into the PCA space, denoted P_{I}, is calculated via:

$$P_{\mathrm{I}} = IC \tag{4}$$

With I now represented in the principal component space as P_{I}, it can be separated into *condition-dependent* and *condition-invariant* dimensions according to a selected parameter p. $\mathrm{P_I}^{\mathrm{dep}}$ is defined as the first p elements of P_{I}. If C is a full decomposition of I; that is, if the number of dimensions in P_{I} is equal to the number of dimensions in I, then $\mathrm{P_I}^{\mathrm{inv}}$ consists of the remaining elements of P_{I}. If C is a not a full decomposition of I then $\mathrm{P_I}^{\mathrm{inv}}$ is calculated via:

$$\mathrm{P_I}^{\mathrm{inv}} = (I - \mathrm{P_I}^{\mathrm{dep}} C^T)C \tag{5}$$

The meaning of the parameter p can be quantified by the proportion of variance within the training dataset that is represented by the condition-dependent information. Assuming a full PCA decomposition, the variance associated with each dimension is equal to the matching eigenvalue of the training set calculated by SVD as described in Equation 3. If the eigenvalues are normalised to 1, the percentage of the variance V_p associated with the condition-dependent information can be calculated as the sum of the first p variances:

$$V_{\mathrm{p}} = \sum_{\mathrm{k}=1}^{\mathrm{p}} e_{\mathrm{k}} \tag{6}$$

The resulting condition-invariant images can then be used as input to a visual place recognition algorithm. Prior place recognition research has shown that removing the low variance components can also improve place recognition on datasets with little appearance change [Liu and Zhang, 2012]. Consequently there might be an intermediate approach that discards both the highest and lowest variance components to provide an optimal set, or one that selects the variance components based on the expected degree of appearance change.

4 Experimental Setup

This section introduces the experiments used to test the PCA learning techniques described above, including the two testing environments used (Section 4.1) and the experiments undertaken (Section 4.2).

4.1 Testing environments

Condition-invariant place recognition was tested on two different datasets. Both datasets have been used in similar place recognition tests [Milford and Wyeth, 2012; Neubert *et al.*, 2013; Sünderhauf *et al.*, 2013], and demonstrate different types of perceptual change. It has already been shown in [Milford and Wyeth, 2012] and [Neubert *et al.*, 2013] that conventional feature-based place recognition algorithms such as FAB-MAP [Cummins and Newman, 2008] are not effective on these changing datasets.

Nordland (winter-spring)

The Nordland dataset is a 700 km long train journey captured by a Norwegian television company over four seasons[1]. This dataset was used for place recognition experiments across changing seasons in [Sünderhauf *et al.*, 2013]. This experiment used images captured in winter and spring because they included some of the most significant perceptual change (see Figure 2 for sample images).

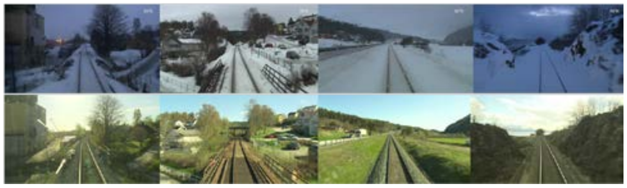


Figure 2. Sample Nordland images from matching locations: winter (top) and spring (bottom).

Data was extracted at 1Hz as in [Sünderhauf *et al.*, 2013] and downsampled to 36 × 64 pixel grayscale-intensity images prior to PCA. After the PCA transformation and the removal of the condition-dependent dimensions, the images were converted back to pixel space and patch-normalised as in the original paper using a patch of size 4. The resulting condition-invariant, patch-normalised images were compared to standard images that had also been patch-normalised using a patch of size 4.

The first 2500 images (from the winter dataset) were used as training data. The number of training images was selected to allow a full PCA decomposition (as each images contained 2048 pixels), although in general full decomposition is not necessary for condition-invariant training (see Section 5.1). The test set started 8 minutes after the training set was completed. This buffer was included to ensure there was a clear break between testing and training sets, and so the ability of the system to generalise to previously unseen locations could be tested.

A subset of the journey of around 80 minutes was used to test the single image matching. When testing with SeqSLAM, the entire train journey (over 8 hours of data and several hundreds of kilometres) was used.

Alderley (sunny day-stormy night)

The Alderley dataset was first presented in [Milford and Wyeth, 2012] and consists of two loops of a suburban street in a car, one during the day, and one at night during a rainstorm. The night images are particularly challenging

[1] http://nrkbeta.no/2013/01/15/nordlandsbanen-minute-by-minute-season-by-season/

due to the darkness, heavy rain and headlight flare which drastically reduce the amount of useful visual information in the images (see Figure 3). For this experiment, the dataset was converted to grayscale and downsampled to a resolution of 24 × 64. After the PCA transformation and the removal of the condition-dependent dimensions, the images were converted back to pixel space and patch-normalised as in the original paper using a patch of size 8. The resulting condition-invariant patch-normalised images were compared to standard images that had also been patch-normalised using a patch of size 8. The system was trained on the first half of the night dataset, and the second half of each loop was used for testing.

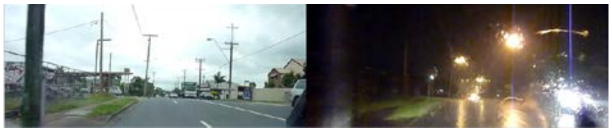


Figure 3. Sample Alderley images from the same location.

4.2 Experiments

The experiments in this paper were designed to test whether place recognition using condition-invariant images outperforms place recognition using standard images. First, examples of condition-invariant images were generated from each dataset to aid with visualisation. Then experiments were performed to (i) test the effect of integrating condition-invariant images into single-image place recognition systems, (ii) test the effect of integrating condition-invariant images into SeqSLAM, and (iii) test the effect of the parameter V_p. This section describes the details of each experiment.

Visualisation

Condition-invariant images were generated to qualitatively illustrate the effect of removing the condition-dependent dimensions from the images. For visualisation ease, larger images were used than for the place matching experiments – the Nordland images were downsampled to 180 × 320 and the Alderley images were downsampled to 130 × 320. The principal component decomposition was learned on the same training set as for the smaller images, and the condition-dependent and condition-invariant aspects were separated based on the parameter value $p = 100$.

Single-image place recognition

This experiment tested how well the system could match places using only appearance information, and without any location prior or sequential matching. With the exception of [Milford *et al.*, 2014], only sequential methods such as SeqSLAM have been shown to perform place matching effectively on highly changing datasets.

For this experiment, two different values of V_p ($V_p = 60\%$ and $V_p = 85\%$) were used for each dataset. These two values provided a preliminary test of the impact of this parameter; in the third experiment, a more comprehensive test was conducted over all possible values of V_p. The patch-normalised images were compared using the L_2 distance, and the minimum distance was used to select the best match for each image. The precision and recall achieved using condition-invariant images were compared to that achieved using standard images.

SeqSLAM

A practical approach to localisation is use a recursive or sequential approach to integrate likelihood over sequences of images. This section presents place recognition results achieved by integrating SeqSLAM with condition-invariant images.

OpenSeqSLAM [Sünderhauf *et al.*, 2013] was used, with the parameters set as in previously published works [Milford and Wyeth, 2012; Sünderhauf *et al.*, 2013]. The place recognition ability of SeqSLAM with condition-invariant images was compared to that of SeqSLAM with standard images. For both datasets, the condition invariance parameter was set to $V_p = 85\%$.

The most significant parameter for SeqSLAM is the sequence length d_s. For the Nordland dataset, SeqSLAM was tested with sequence lengths of 3, 5, 10 and 20 images, which is equivalent to distances of approximately 90 m, 150 m, 300 m and 600 m (assuming a typical speed of 30 ms^{-1}). The Alderley dataset was tested with sequence lengths of 25 and 50 images, representing typical distance lengths of around 25 m and 50 m, significantly shorter than the ~300 metre sequence length used in the original study [Milford and Wyeth, 2012].

Parameter choice

The third experiment was an analysis of the effect of V_p on matching capability. As discussed in Section 3.2, the value of V_p relates to the proportion of variance within the training dataset that is represented by the global conditions. The single-image place recognition experiment was repeated on both the Alderley and Nordland datasets across all values of V_p to compare the relationship between variance and localisation performance across different environments and types of change. The performance metric chosen was the maximum recall at 95% precision measured against the percentage of variance considered to be due to the condition-dependent dimensions. The value of 95% precision (rather than 99% or 100% precision) was selected to represent the general behaviour of the dataset, as the higher precision levels are more susceptible to noise.

5 Results

This section presents results for the condition-invariant image experiments described in Section 4.2. The section begins with some examples of condition-invariant images from each dataset, showing the effect of the condition-invariant training on different environmental conditions at the same location. Precision-recall curves are used to present the results from the single-image place recognition and sequence-based place recognition experiments. Finally, the parameter V_p is plotted against maximum recall at 95% precision.

5.1 Visualisation

Figure 4 shows the effect of condition-invariant training on an example location from the Nordland dataset and Figure 1 shows example locations from the Alderley dataset. From a qualitative perspective, the condition-invariant images appear more similar than the original images. For example, in the Nordland example the snow-covered areas have been “faded out”, while in the Alderley examples, much of the intensity difference between sky and headlight/street light locations has been removed. These condition-invariant images give a qualitative impression of what the process is doing, while the following sections detail the quantitative place recognition performance.

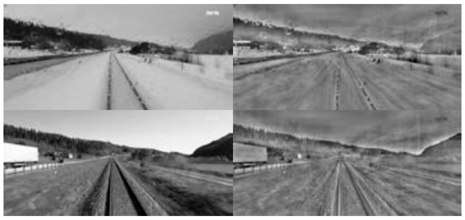


Figure 4. Sample location from the Nordland dataset displaying condition-invariant images (right column) and original images (left column). The condition-invariant images appear more similar across different seasons than the original images.

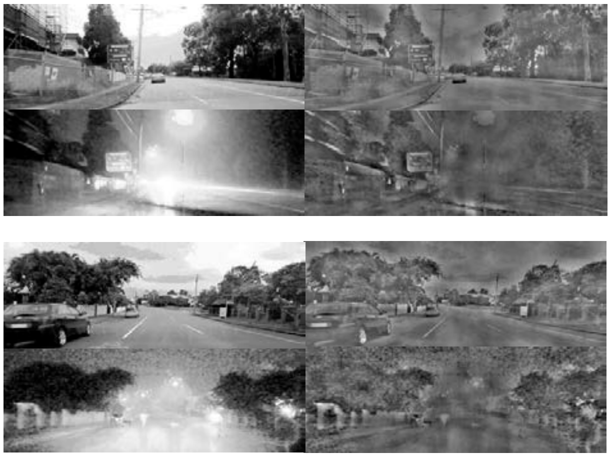


Figure 5. Sample locations from the Alderley dataset displaying condition-invariant images (right column) and original images (left column). The condition-invariant images appear more similar across different conditions than the original images.

5.2 Single-image place recognition

Precision-recall curves for single-image place matching on the Nordland and Alderley datasets are shown in Figure 6 and Figure 7 respectively. Performance using standard patch-normalised images is shown by the red dashed line, with the performance using condition-invariant patch normalised images shown in black (for the parameter values V_p = 60% and V_p = 85% for each dataset).

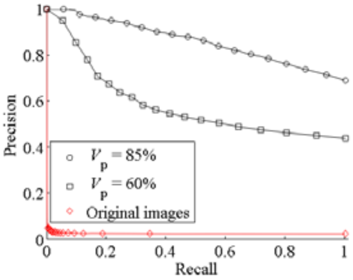


Figure 6. Precision-recall curves for the Nordland dataset with condition-invariant (black) and untrained (red) images. Recall at 100% precision is 0.06% for untrained images, 1.7% for V_p = 60% and 7.6% for V_p = 85%.

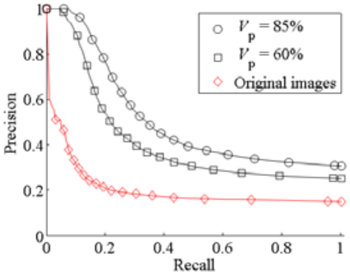


Figure 7. Precision-recall curves for the Alderley dataset with condition-invariant (black) and untrained (red) images Recall at 100% precision is 0.03% for untrained images, 4.7% for V_p = 60% and 7.6% for V_p = 85%.

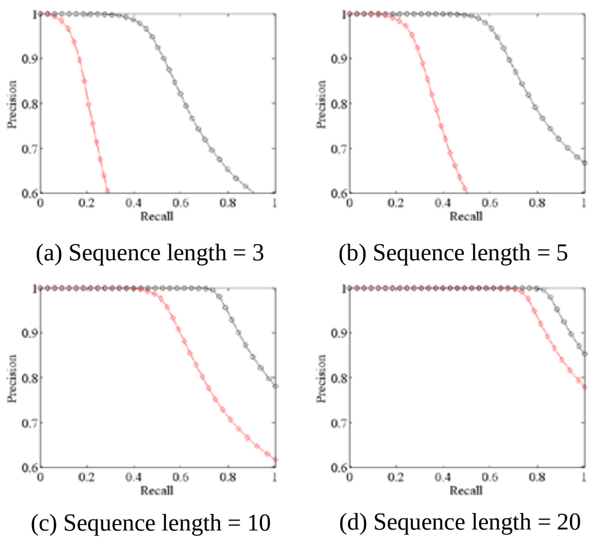
For both datasets, performance improves significantly when using condition-invariant images. Standard single-image matching demonstrated no practical place recognition capability on either dataset (0.06% and 0.03% recall at 100% precision respectively), demonstrating how challenging appearance-based matching is on these datasets. Using condition-invariant images, recall at 100% precision is 7.6% for both datasets when V_p = 85%.

The results presented here are competitive with other single-image place recognition techniques that benefit from more laborious training schemes. On the Nordland dataset, SC-ACP [Neubert *et al.*, 2013] achieved recall of only around 2% at 100% precision on a 22 minute subset, and the recall curve decayed quickly – for example, at 90% precision the recall had only increased to 5% compared to an increase to over 40% for the condition-invariant system here.

On the Alderley dataset, [Milford *et al.*, 2014] achieved higher recall at 100% precision (21.2%) using a patch verification technique on high resolution images. However, the first step in this process uses whole-image matching to generate hypotheses for further verification, and the condition-invariant system significantly out-performs this step, which never achieves 100% precision. A potential integration between the condition-invariant system and a multi-hypothesis verification stage could be valuable, with the condition-invariant images used to generate high quality hypotheses and the verification stage using high resolution data to accept or reject the results.

5.3 Sequence-based place recognition

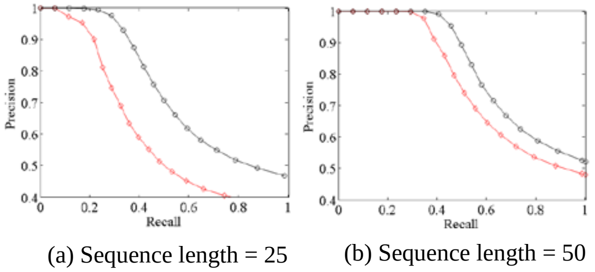
The precision-recall curves for SeqSLAM with standard images and SeqSLAM with condition-invariant images are shown for the Nordland dataset in Figure 8. The condition-invariant version displays a significant improvement over the standard input. The condition-invariant images provide reasonable localisation performance with sequences as short as 3 images, achieving 16.1% recall at 100% precision, more than 4 times the recall for the standard images (3.8% recall at 100% precision). Maintaining localisation performance with a reduced sequence length is of practical importance since shorter sequence lengths allow greater search efficiency, lower latencies and enable identification of shorter contiguous road segments [Pepperell *et al.*, 2013].



(a) Sequence length = 3
(b) Sequence length = 5
(c) Sequence length = 10
(d) Sequence length = 20

Figure 8. Precision-recall curves on Nordland dataset using SeqSLAM with condition-invariant (black) and untrained (red) images, and with sequence lengths of (a) 3, (b) 5, (c) 10 and (d) 20. Condition-invariant images outperform untrained images in both cases.

The precision-recall curves for SeqSLAM with standard images and SeqSLAM with condition-invariant images are shown for the Alderley dataset in Figure 9. Using a 25 image sequence the recall at 100% precision nearly doubles from 8.1% to 15.8%.



(a) Sequence length = 25
(b) Sequence length = 50

Figure 9. Precision-recall curves on Alderley dataset using SeqSLAM with condition-invariant (black) and untrained (red) images, and with sequence lengths of (a) 25 and (b) 50. Condition-invariant images outperform untrained images in both cases.

5.4 Parameter choice

The results from Sections 5.2 and 5.3 demonstrate that removing the condition-dependent dimensions from an image set can improve place recognition performance across changing environmental conditions. This section examines the relationship between the proportion of the training set variance that is associated to the condition-dependent dimensions (via the choice of parameter V_p) and the localisation performance of the condition-invariant comparison for each dataset.

The relationship between the percentage of variance V_p assigned to the condition-dependent dimensions against the maximum recall at 95% precision is plotted in Figure 10 and Figure 11 for the Nordland and Alderley datasets respectively. The results are displayed at 95% precision, rather than at 100% precision, as the results are less noisy; at 100% precision there is more sensitivity to slight changes in V_p and it is difficult to determine the relationship between V_p and recall. Although these plots only show the results at 95% precision, a similar trend is observed at other precision levels as well.

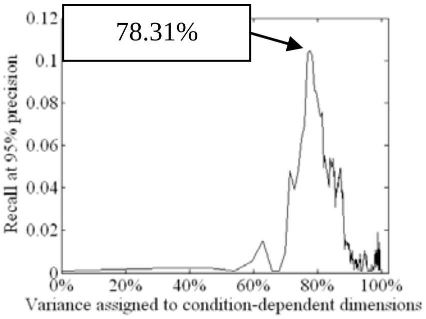


Figure 10. Variance assigned to condition-dependent dimensions (V_p) against maximum recall at 95% precision for the Nordland dataset. Recall at 95% precision peaked when 78.31% of variance was removed from the images.

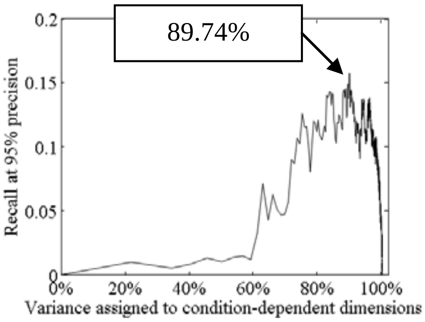


Figure 11. Variance assigned to condition-dependent dimensions (V_p) against recall at 95% precision for the Alderley dataset. Recall at 95% precision peaked when 89.74% of variance was removed from the images.

The datasets display qualitative similarities in terms of the effect of V_p on the behaviour of the system. The best performance is achieved when close to 80% (for the Nordland dataset) and 90% (for the Alderley dataset) of the variance is assigned to condition-dependent dimensions, and only the final 20% or 10% is retained for place recognition matching. Note that V_p is calculated on the training data but precision and recall were measured on the test set. Nonetheless, removing 80% (or more) of the principal component variance can have a dramatic effect on place recognition, demonstrating that the learned condition-invariance has generalised to new spatial locations.

6 Discussion

This paper presents a method that improves place matching by removing elements from the environment that are not useful for matching. The method is easy to implement, using only conventional principal component analysis. Its novelty in a place recognition context comes from its inversion of the typical application of PCA; the most significant principal components are removed and place recognition is performed on the remaining components. The experiments demonstrate that negative effects of seasonal, lighting and weather changes on place recognition performance are reduced by removing the early dimensions from the PCA decomposition. Furthermore, this decomposition can be learned without knowing frame correspondences between observations from identical locations at different times, making the training process significantly more practical than existing learning-based schemes.

There is still an open question of how the parameter V_p, the value which defines where the condition-dependent and condition-invariant dimensions are partitioned, can be determined in a principled manner. There are a number of methods for automatically selecting the number of principal components [Jolliffe, 2002]. More sophisticated matrix factorisation methods such as Principal Component Pursuit [Candès *et al.*, 2011] may be feasible, whilst techniques such as deep learning also offer dimensionality reduction techniques that are superior to PCA [Hinton and Salakhutdinov, 2006] and could potentially be applied to the place recognition problem.

The condition-invariant information could also be used in more sophisticated ways. The strength of the system lies in its ability to remove what is there. It does not infer what *should* be there, and this is particularly apparent in the Alderley dataset, where the occlusion caused by headlight glare is successfully removed, but the parts of the image occluded by the headlights lack any information (Figure 1). The existing system could be enhanced by using these results to identify and mask such information-deprived areas during the later image comparison phases.

A related question is how this approach might integrate with other techniques for place recognition, mapping and SLAM over long periods of time and changing conditions. This paper does not address the pose invariance problem, but the condition-invariant learning method presented here is not itself pose-dependent, enabling the future possibility of integration into more pose invariant place recognition techniques.

Acknowledgements

We thank Andrew English for useful comments on early versions of this work. This work was in part funded by the Australian Research Council through DECRA grant DE120100995 to MM. SL is supported by an Australian Postgraduate Award.

References

[Artac *et al.*, 2002] M. Artac, M. Jogan and A. Leonardis, "Mobile robot localization using an incremental eigenspace model," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, 2002, pp. 1025-1030.

[Badino *et al.*, 2012] H. Badino, D. Huber and T. Kanade, "Real-time topometric localization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1635-1642.

[Bay *et al.*, 2008] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, pp. 346-359, Jun 2008.

[Cadena and Neira, 2011] C. Cadena and J. Neira, "A learning algorithm for place recognition," presented at the ICRA 2011 Workshop on Long-term Autonomy, Shanghai, China, 2011.

[Candès *et al.*, 2011] E. J. Candès, X. Li, Y. Ma and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, p. 11, 2011.

[Carlevaris-Bianco and Eustice, 2014] N. Carlevaris-Bianco and R. M. Eustice, "Learning Visual Feature Descriptors for Dynamic Lighting Conditions," presented at the Proc. of Workshop on Visual Place Recognition in Changing Environments, IEEE International Conference on Robotics and Automation (ICRA), 2014.

[Carreira *et al.*, 2014] F. Carreira, J. F. Calado, C. Cardeira and P. Oliveira, "Enhanced PCA-Based Localization Using Depth Maps with Missing Data," *Journal of Intelligent & Robotic Systems*, pp. 1-20, 2014/01/24 2014.

[Chandrasekaran *et al.*, 2011] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, pp. 572-596, 2011.

[Corke *et al.*, 2013] P. Corke, R. Paul, W. Churchill and P. Newman, "Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, pp. 2085-2092.

[Cummins and Newman, 2011] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *International Journal of Robotics Research*, vol. 30, pp. 1100-1123, 2011.

[Cummins and Newman, 2008] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *International Journal of Robotics Research*, vol. 27, pp. 647-665, Jun 2008.

[Dudek and Jugessur, 2000] G. Dudek and D. Jugessur, "Robust place recognition using local appearance based methods," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, 2000, pp. 1030-1035.

[Hinton and Salakhutdinov, 2006] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504-507, 2006.

[Hotelling, 1933] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, pp. 417-441,498-520, 1933.

[Johns and Yang, 2013] E. Johns and G.-Z. Yang, "Feature co-occurrence maps: Appearance-based localisation throughout the day," in *Proc. ICRA*, 2013.

[Jolliffe, 2002] I. Jolliffe, *Principal Component Analysis*, 2nd ed.: Springer, 2002.

[Liu and Zhang, 2012] Y. Liu and H. Zhang, "Visual loop closure detection with a compact image descriptor," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 1051-1056.

[Lowry *et al.*, 2014] S. Lowry, M. Milford and G. Wyeth, "Transforming morning to afternoon using linear regression techniques," in *IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 2014.

[Maddern *et al.*, 2014] W. Maddern, A. D. Stewart, C. McManus, B. Upcroft, W. Churchill and P. Newman, "Illumination Invariant Imaging: Applications in Robust Vision-based Localisation, Mapping and Classification for Autonomous Vehicles," *Proc. of Workshop on Visual Place Recognition in Changing Environments, IEEE International Conference on Robotics and Automation (ICRA)*, 2014.

[McManus *et al.*, 2013] C. McManus, P. Furgale and T. D. Barfoot, "Towards lighting-invariant visual navigation: An appearance-based approach using scanning laser-rangefinders," *Robotics and Autonomous Systems*, 2013.

[McManus *et al.*, 2014] C. McManus, B. Upcroft and P. Newman, "Scene Signatures: Localised and Point-less Features for Localisation," in *Robotics Science and Systems*, 2014.

[Milford *et al.*, 2014] M. Milford, W. Scheirer, E. Vig, A. Glover, O. Baumann, J. Mattingley and D. Cox, "Condition-Invariant, Top-Down Visual Place Recognition," in *IEEE International Conference in Robotics and Automation (ICRA)*, 2014.

[Milford and Wyeth, 2012] M. Milford and G. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1643-1649.

[Naseer *et al.*, 2014] T. Naseer, L. Spinello, W. Burgard and C. Stachniss, "Robust visual robot localization across seasons using network flows," in *Conference on the Association for the Advancement of Artificial Intelligence*, 2014.

[Neubert *et al.*, 2013] P. Neubert, N. Sünderhauf and P. Protzel, "Appearance change prediction for long-term navigation across seasons," in *European Conference on Mobile Robots (ECMR)*, 2013.

[Pearson, 1901] K. Pearson, "On lines and planes of closest fit to systems of points in space.," *Philosophical Magazine*, vol. 6, pp. 559-572, 1901.

[Pepperell *et al.*, 2013] E. Pepperell, P. Corke and M. Milford, "Towards persistent visual navigation using SMART," presented at the Proceedings of Australasian Conference on Robotics and Automation, University of New South Wales, Sydney, Australia, 2013.

[Ranganathan *et al.*, 2013] A. Ranganathan, S. Matsumoto and D. Ilstrup, "Towards illumination invariance for visual localization," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, 2013, pp. 3791-3798.

[Sirovich and Kirby, 1987] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *JOSA A*, vol. 4, pp. 519-524, 1987.

[Sünderhauf *et al.*, 2013] N. Sünderhauf, P. Neubert and P. Protzel, "Are we there yet? Challenging SeqSLAM on a 3000 km journey across all four seasons," in *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[Turk and Pentland, 1991] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71-86, 1991.

[Valgren and Lilienthal, 2010] C. Valgren and A. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, pp. 157-165, Feb 2010.