Life-Long Place Recognition by Shared Representative Appearance Learning

Fei Han¹, Xue Yang¹, Yiming Deng², Mark Rentschler³, Dejun Yang¹, and Hao Zhang¹

1. Division of Computer Science, Colorado School of Mines
2. Department of Electrical Engineering, University of Colorado Denver
3. Department of Mechanical Engineering, University of Colorado Boulder
This work addresses long-term place recognition with strong appearance variations.
Introduction

The Problem

• Which feature is more representative while insensitive to appearance changes?

• How to fuse these heterogeneous feature modalities?
Contributions

*Shared Representative Appearance Learning (SRAL) model for long-term place recognition*

**Capabilities**

- Autonomous discriminative feature modality learning and fusion that is robust to strong appearance variations
- Sparse optimization formulation regularized by structured sparsity-inducing norms
- Efficient solver with theoretical convergence guarantee
Approach
Scene Representation

• A set of scenes is represented by

\[ X = \begin{bmatrix} x_1 & \cdots & x_i & \cdots & x_n \end{bmatrix} \in \mathbb{R}^{d \times n} \]

\[ x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{im} \end{bmatrix} \in \mathbb{R}^{d} \]

Feature modality 1

Feature modality m

• Scene scenarios are represented as

\[ Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^{n \times c} \]
Approach
Shared Representative Appearance Learning

Formulation: \[ \min_{\mathbf{W}} \|\mathbf{X}^\top \mathbf{W} - \mathbf{Y}\|^2_F + \lambda_1 \|\mathbf{W}\|_{2,1} + \lambda_2 \|\mathbf{W}\|_M \]

\(\ell_{2,1}\)-norm: enforces sparsity of each row of \(\mathbf{W}\)

\(\ell_M\)-norm: enforces sparsity of each feature modality
Approach
Optimization Algorithm

Algorithm: An efficient algorithm to solve the optimization problem

Input: \( X = [x_1, \ldots, x_n] \in \mathbb{R}^{d \times n} \) and
\( Y = [y_1, \ldots, y_n]^\top \in \mathbb{R}^{n \times c} \)

1. Let \( t = 1 \). Initialize \( W(t) \) by solving
\[ \min_W \| X^\top W - Y \|_F^2. \]
2. while not converge do
3. Calculate the block diagonal matrix \( D(t + 1) \), where the \( i \)-th diagonal element of \( D(t + 1) \) is \( \frac{1}{2\|w_i(t)\|_2} I_i \).
4. Calculate the block diagonal matrix \( \tilde{D}(t + 1) \), where the \( i \)-th diagonal block of \( \tilde{D}(t + 1) \) is \( \frac{1}{2\|w_i(t)\|_F} I_i \).
4. For each \( w_i(1 \leq i \leq c), w_i(t + 1) = \left( XX^\top + \gamma_1 D(t + 1) + \gamma_2 \tilde{D}(t + 1) \right)^{-1} X y_i. \)
5. \( t = t + 1. \)
6. end
Output: \( W = W(t) \in \mathbb{R}^{d \times c} \)

Theorem:
The Algorithm converges to the global optimal solution of the formulated optimization problem
Approach
Place Recognition

• Optimal weight $\omega_i$ for each feature modality $i$, $i = 1, ..., m$

$$\omega_i = \| W^{i*} \|_F$$

• Feature fusion for image matching

$$s(q, p) = \sum_{i=1}^{m} \overline{\omega}_i s_i (q, p)$$
Experiments

- Three public datasets
  - Different time of the day: St Lucia Dataset
  - Different months: CMU-VL Dataset
  - Different seasons: Nordland Dataset

- Setups
  - Hardware: Workstation with i7 3.4GHz, 16GB
  - Feature modalities: Color, GIST, HOG, and LBP features
Experiments
St Lucia Dataset (Various Times of the Day)

Qualitative Evaluation
Experiments
St Lucia Dataset (Various Times of the Day)

Quantitative Evaluation

![Graph showing precision-recall curve with various features like BRIEF-GIST, Feature Concatenation, SRAL, SeqSLAM, Color, and LBP.]

![Pie chart showing contribution percentages: GIST 40%, HOG 59.1%, Color 0.6%, LBP 0.3%]
Experiments
CMU-VL Dataset (Different Months)

Qualitative Evaluation
Experiments
CMU-VL Dataset (Different Months)

Quantitative Evaluation

- BRIEF-GIST
- Feature Concatenation
- SRAL
- SeqSLAM
- Color
- LBP

Precision vs. Recall graph

- HOG 88.25%
- GIST 0.98%
- Color 9.79%
- LBP 0.98%
Experiments
Nordland Dataset (Different Seasons)

Qualitative Evaluation
Experiments
Nordland Dataset (Different Seasons)

Quantitative Evaluation

- BRIEF-GIST
- Feature Concatenation
- SRAL
- SeqSLAM
- Color
- LBP

Precision vs Recall

Color 2.49%
GIST 0.2%
LBP 0.64%

HOG 96.68%
We introduced a long-term place recognition method, named *Shared Representative Appearance Learning (SRAL)*, which

- Can autonomously learn and integrate shared discriminative features that are robust to strong appearance variations
- Formulates the task as a convex optimization problem regularized by structured sparsity
- Can be efficiently solved with theoretical convergence guarantee

**Life-Long Place Recognition by Shared Representative Appearance Learning**