Development of Digital Scene Matching Area Correlation (DSMAC) Algorithms and Prototype System

PhD Scholar: Anuradha Uggi, PI: Sumohana S. Channappayya, Co-PI: Amit Acharyya

Indian Institute of Technology Hyderabad

Motivation

- Don't know the way to the destination?
  - No problem, we have GPS. What if the GPS is inaccessible?
- Military: Areas where spoofing, jammers, or signal blockers are installed
- Civil applications
  - Indoor Surveillance
  - Search and rescue
  - Disaster zones
  - Essentials delivery in unreachable places
  - Data collection, etc.
- Visual Place Recognition (VPR) as an alternative to GPS

Basic Set-up of Visual Place Recognition

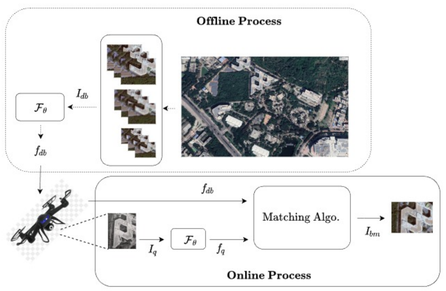

Figure 1. Pipeline of the VPR set-up

Practical Challenges

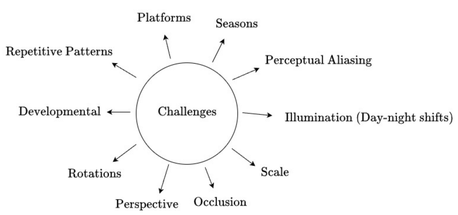

Figure 2. Practical Challenges

- Rotation and perspective changes: cluster-based techniques
  - NetVLAD [Arandjelovic et al. CVPR 2016]
  - Patch-NetVLAD [Hausler et al. CVPR 2021]
  - MixVPR [Ali-bey et al. WACV 2023] [Limited invariance to rotation changes]
  - **MS-NetVLAD** [Uggi and Channappayya SPL 2024]
- Scale changes: Multi-scale information during training
  - MR-NetVLAD [Khaliq et al. RAL 2022]
- Day-Night shifts or illumination changes: Multi-sensor information
  - **Analysis of RGB pretrained models on RGB-Thermal Retrieval** [Pendota and Channappayya CVPRW 2024]
- Perceptual Aliasing
  - **Sequence matching** [Gupta IITH MTech Thesis 2024]

Illustrations for Challenging Scenarios

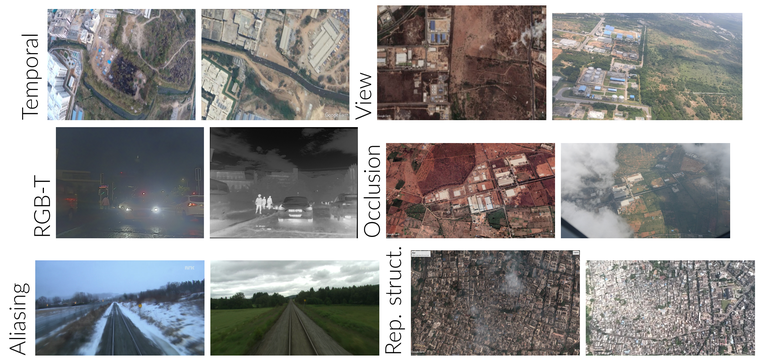

Figure 3. Sample images in different scenarios

Analysis and Contributions

- **NetVLAD** [Arandjelovic et al. CVPR 2016]: Cluster the features and use cluster centres as proxies for retrieval
- **MixVPR** [Ali-bey et al. WACV 2023]: Mix the features
- **MS-NetVLAD** [Uggi and Channappayya SPL 2024]: Is the aggregation of features across intermediate layers benefits over the bottleneck features?
- **Pretrained Models on the RGB-T Datasets** [Pendota and Channappayya CVPRW 2024]: Are off-the-shelf multi-purpose models helpful for RGB-T retrieval?

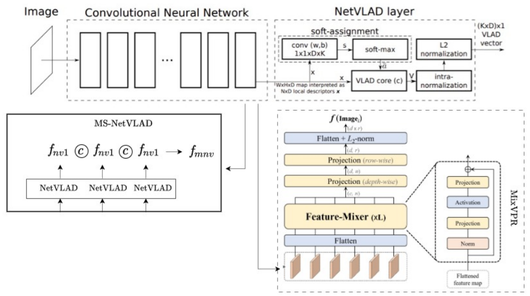

Figure 4. A unified representation of the NetVLAD, MixVPR, and MS-NetVLAD

Quantitative Results

Dataset	Metric	NetVLAD	Patch-NetVLAD	MixVPR
DTVA w/ N16	R@1	38.02	67.09	**72.24**
	R@5	53.66	78.25	**81.88**
DTVA w/ N16 (scaled)	R@1	22.43	46.68	**58.40**
	R@5	39.7	67.01	**70.72**
DTVA w/ N16 (rotated)	R@1	22.39	**48.49**	38.88
	R@5	39.78	**68.37**	46.52

Table 1. Results on real-time queries vs satellite database

Qualitative Results

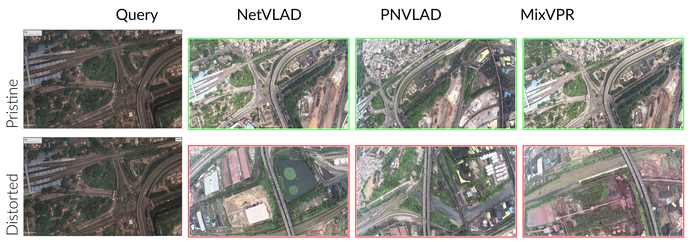

Figure 5. Results of off-the-shelf models

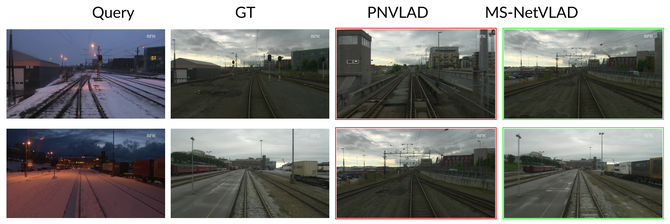

Figure 6. Results of MS-NetVLAD

Figure 7. Results of pretrained models on RGB-T datasets

Testing on Hardware

- MixVPR runs at a speed of 20 fps on the DICE board, a DRDO-designed hardware and 30 fps on the Jetson Orin nano hardware

Conclusion

- Extensive analysis is done to identify the model (MixVPR) that best suits the real-time requirements
- Accuracy numbers alone are not enough to trust the AI models. Specifically in risk-sensitive applications
- We aim to introduce Explainability in our future work