VII. Supplementary material

A. PR AUC and Recall@1 Per Sequence Pairing

To provide more insights into how the various methods reported in Table III perform on each of the sequence pairings, Fig. 5 shows the precision-recall curves resulting from exhaustive pairwise matching of each sequence combination. Since Multi-Process Fusion does not provide an interface for exhaustive matching, we are not able to add it to the precision-recall plots. We further report retrieval performance for $N = 1$ on each combination of query and database sequences in Table VII. As outlined in Section V-B, we compare our composite descriptor to both visual and structural descriptors. We further evaluate against NetVLAD variants obtained using either off-the-shelf (ots) or fine-tuned (ft) weights as well as against Multi-Process Fusion, DenseVLAD and SeqSLAM for sequence lengths of $20\,\mathrm{m}$ and $40\,\mathrm{m}$, respectively. For this comparison, all descriptors including those produced by NetVLAD and DenseVLAD are projected to $dim_f = 256$ dimensions. Our composite descriptor outperforms the single-modality descriptors as well as all of the baselines in exhaustive pairwise matching of each sequence combination. On top of that, it shows best retrieval performance on all except two sequence pairings. Our results further reveal that fine-tuning of NetVLAD on our training dataset somewhat deteriorates the pairwise matching accuracy for several sequence combinations but improves retrieval performance for the majority of sequence pairings. Similar observations have been reported in [23].

B. Network Configurations

TABLE V

The number of output channels and the arrangement of pooling operations for each of the d_S layers of the structure-based feature extraction networks used to obtain the results reported in Table II.

d_S	# Channels per CONV-layer	Pooling indices
6	$2 \times [32], 2 \times [64], 2 \times [128]$	$2, 3, 4$
8	$2 \times [32], 3 \times [64], 3 \times [128]$	$2, 4, 6$
9	$2 \times [32], 4 \times [64], 3 \times [128]$	$2, 4, 6, 8$
10	$2 \times [32], 4 \times [64], 4 \times [128]$	$2, 4, 6, 8$
12	$2 \times [32], 5 \times [64], 5 \times [128]$	$2, 4, 7, 10$

In Table V, we provide an overview of all structure-based feature extraction network configurations evaluated and compared in Section V-A. The pooling indices denote which convolutional layers are followed by pooling operations – with the networks with depths $d_S = 6$ and $d_S = 12$ not following the pattern described in Section III-A. Furthermore, Table VI shows the feature extraction configurations used in our best performing compound network as well as the slightly modified versions used to obtain the 256-dimensional appearance and structural descriptors reported in Table III.

C. Image Retrieval

We provide additional examples of top retrieved database instances when using either appearance, structure, or composite descriptors, see Fig. 6. Composite descriptors are produced using the best-performing model architecture which uses $d_A = 12$ and $d_S = 9$ layers for both visual and structural feature extraction networks, respectively, and a simple concatenation of the extracted intermediary features. While descriptors that only encode either appearance or structure often produce incorrect matches, each of the shown queries is correctly matched when using our composite descriptor.

TABLE VI

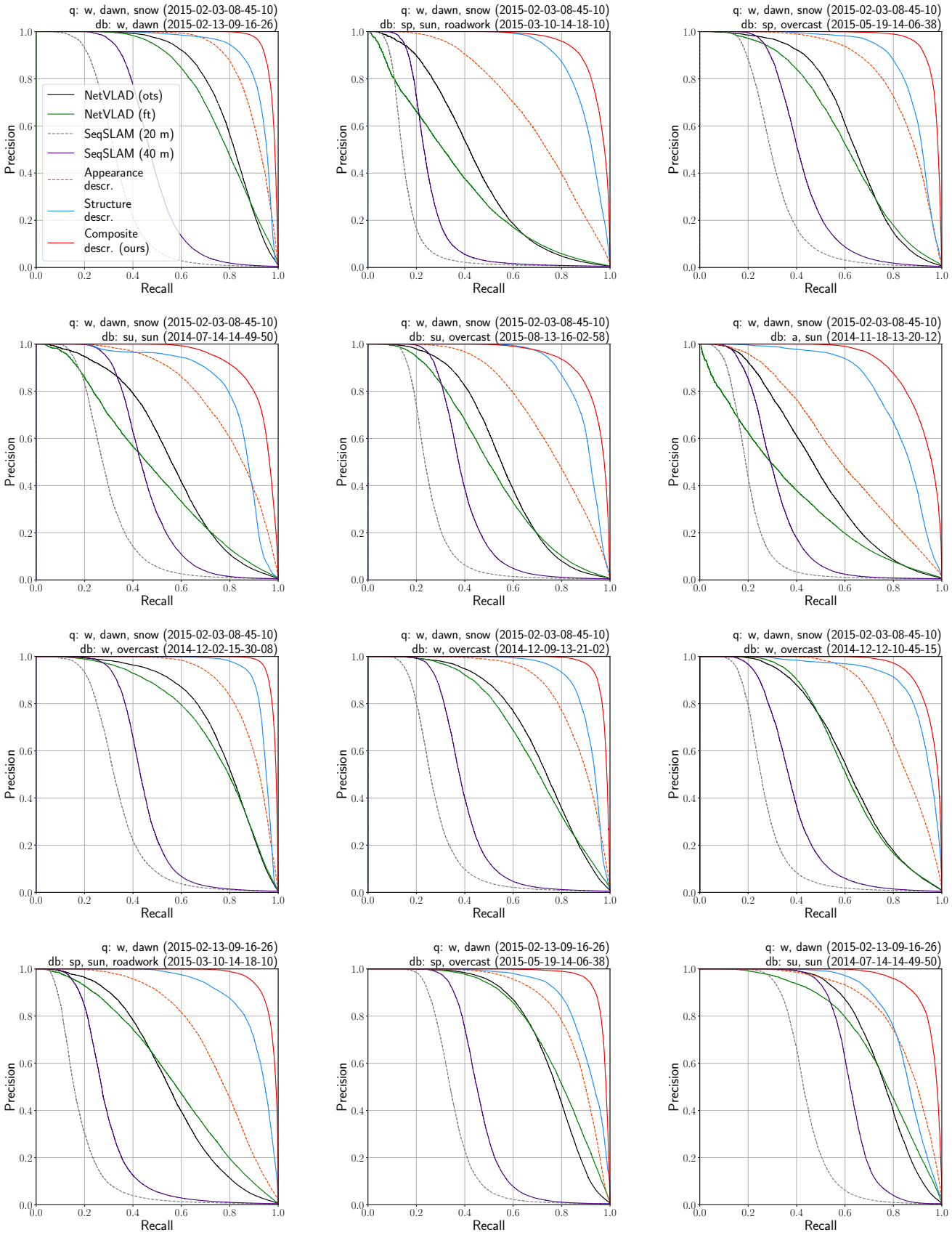
Overview of the apperance- (A) and structure-based (S) feature extraction network configurations reported in Table III. Listed are the configuration of our best performing compound network as well as the slightly modified versions used to generate the 256-dimensional single-modality descriptors.

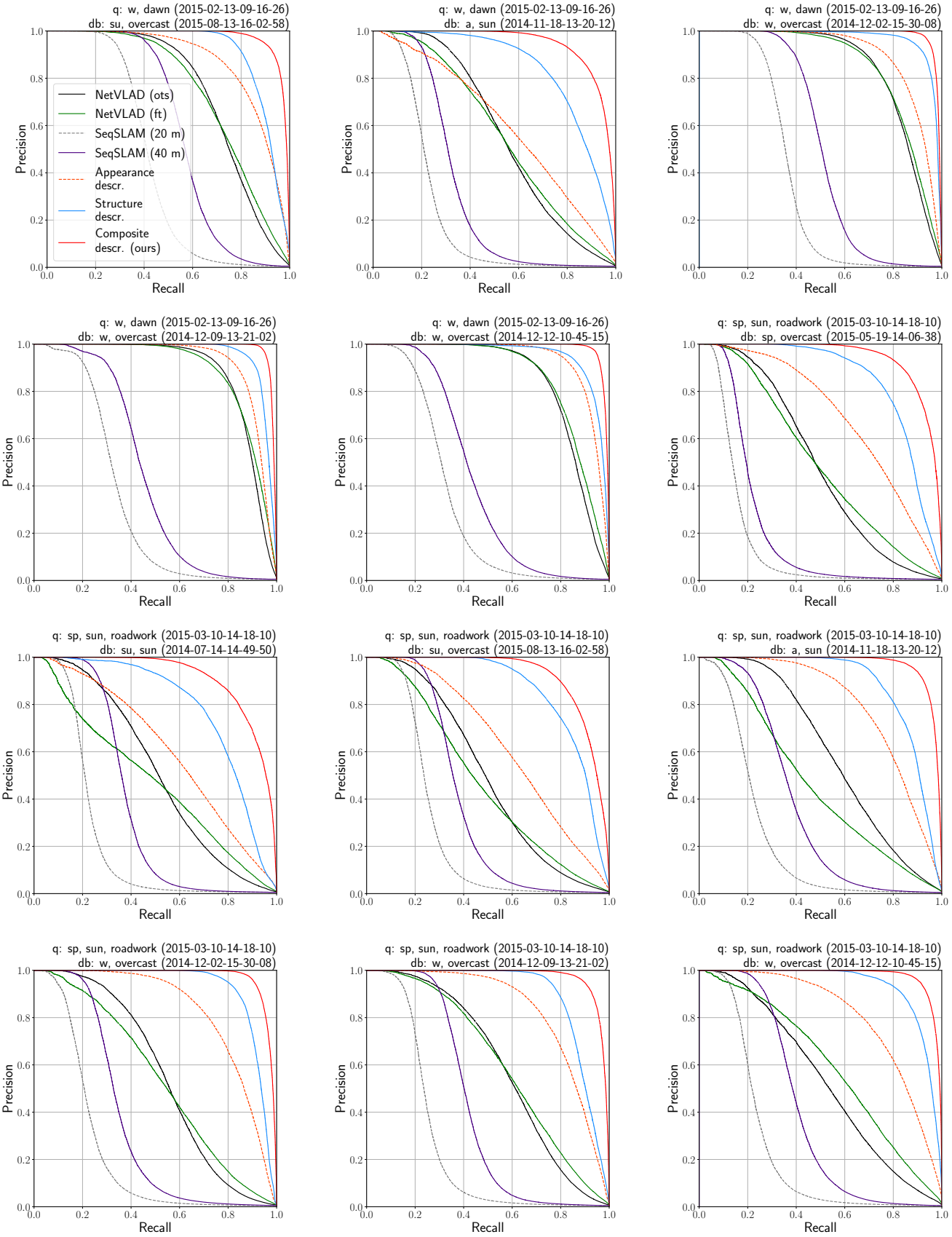
Type	d	# Channels per CONV-layer	Pooling indices
Compound network			
A	12	$6 \times [64], 6 \times [128]$	$2, 4, 6, 8, 10$
S	9	$2 \times [32], 4 \times [64], 3 \times [128]$	$2, 4, 6, 8$
Appearance- / structure-based networks			
A	12	$5 \times [64], 5 \times [128], 2 \times [256]$	$2, 4, 6, 8, 10$
S	9	$2 \times [32], 3 \times [64], 2 \times [128], 2 \times [256]$	$2, 4, 6, 8$

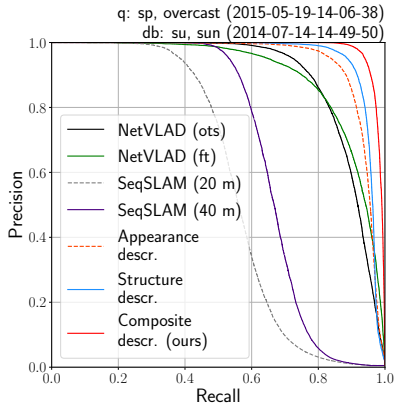
TABLE VII

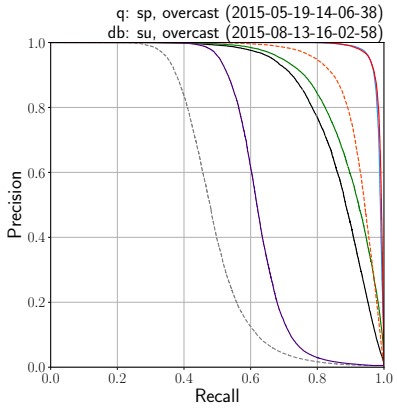
RECALL FOR $N = 1$ FOR EACH PAIRING OF ROBOTCAR SEQUENCES. A QUERY LOCATION IS DEEMED CORRECTLY MATCHED IF THE RETRIEVED DATABASE INSTANCE EXHIBITS A GROUND TRUTH DISTANCE OF LESS THAN 20 m. EACH SEQUENCE IS DENOTED USING ITS ABBREVIATED TIMESTAMP (*mmddss*). WE UNDERLINE THE BETTER VALUES WHEN COMPARING OFF-THE-SHELF AND FINE-TUNED NETVLAD DESCRIPTORS AS WELL AS APPEARANCE- AND STRUCTURE-BASED DESCRIPTORS. FINE-TUNING NETVLAD IMPROVES ITS RETRIEVAL PERFORMANCE FOR THE MAJORITY OF SEQUENCE PAIRINGS. SUPERIORITY OF USING EITHER VISUAL OR STRUCTURAL CUES VARIES ACROSS SEQUENCE PAIRINGS. HOWEVER, FUSING BOTH MODALITIES INTO OUR COMPOSITE REPRESENTATION RESULTS IN BEST RETRIEVAL PERFORMANCE (PRINTED IN BOLD) ON ALL EXCEPT TWO SEQUENCE PAIRINGS. THE AVERAGES ACROSS ALL SEQUENCE PAIRINGS ARE ALSO REPORTED IN TABLE III.

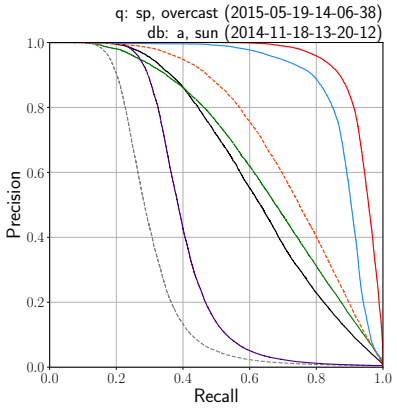
Query seq.	Database seq.	NetVLAD (ots)	NetVLAD (ft)	MPF	DenseVLAD	SeqSLAM (20 m)	SeqSLAM (40 m)	Appearance descr.	Structural descr.	Composite descr.
020310	**021326**	0.9669	0.9722	0.9823	0.9678	0.7555	0.8312	0.9751	0.9861	**0.9890**
020310	**031010**	0.7355	0.6743	0.7905	0.4845	0.4675	0.5471	0.9780	0.9204	**0.9906**
020310	**051938**	0.8584	0.8792	0.9156	0.7902	0.6733	0.7821	0.9306	0.9845	**0.9955**
020310	**071450**	0.8438	0.8072	0.7329	0.6287	0.6943	0.7929	0.8744	0.9542	**0.9786**
020310	**081358**	0.8473	0.8269	0.9053	0.7478	0.6162	0.6731	0.9302	0.9514	**0.9816**
020310	**111812**	0.7518	0.7241	0.7901	0.5743	0.4810	0.6031	0.9086	0.9008	**0.9776**
020310	**120208**	0.9359	0.9433	0.9626	0.8857	0.7173	0.7715	0.9792	0.9890	**1.0000**
020310	**120902**	0.9249	0.9294	0.9354	0.8849	0.6844	0.7736	0.9755	0.9820	**0.9996**
020310	**121215**	0.8429	0.8312	0.8255	0.7886	0.6830	0.7297	0.9161	0.9502	**0.9583**
021326	**031010**	0.7984	0.8126	0.8060	0.7293	0.4559	0.5784	0.9792	0.9146	**0.9875**
021326	**051938**	0.9159	0.9534	0.9744	0.9584	0.6995	0.8100	0.9525	0.9833	**0.9992**
021326	**071450**	0.8948	0.9178	0.7530	0.8864	0.8012	0.8988	0.8645	0.9492	**0.9744**
021326	**081358**	0.9159	0.9267	0.9492	0.9279	0.7246	0.8290	0.9559	0.9575	**0.9863**
021326	**111812**	0.8247	0.8555	0.8589	0.8442	0.5681	0.6667	0.9413	0.9038	**0.9775**
021326	**120208**	0.9721	0.9733	0.9933	0.9842	0.7598	0.8309	0.9988	0.9888	**1.0000**
021326	**120902**	0.9867	0.9858	0.9891	0.9913	0.6587	0.7449	0.9896	0.9925	**1.0000**
021326	**121215**	0.9012	0.9233	0.8513	0.9007	0.6603	0.6989	0.9122	0.9370	**0.9476**
031010	**051938**	0.8063	0.8687	0.8738	0.6218	0.4594	0.5517	0.9219	0.9273	**0.9788**
031010	**071450**	0.7808	0.8337	0.6978	0.5452	0.5205	0.6508	0.8887	0.8462	**0.9705**
031010	**081358**	0.7984	0.8466	0.8529	0.5777	0.5888	0.7161	0.9273	0.8807	**0.9755**
031010	**111812**	0.9252	0.9381	0.9455	0.8433	0.5450	0.6957	0.9472	0.9443	**0.9875**
031010	**120208**	0.8246	0.8865	0.9790	0.7510	0.5691	0.6896	0.9850	0.9622	**1.0000**
031010	**120902**	0.8263	0.8990	0.9547	0.7469	0.5858	0.6848	0.9618	0.9414	**1.0000**
031010	**121215**	0.8075	0.8837	0.8738	0.6839	0.6063	0.7342	0.9297	0.9110	**0.9521**
051938	**071450**	0.9579	0.9636	0.7674	0.9602	0.8697	0.8905	0.9411	0.9717	**0.9931**
051938	**081358**	0.9674	0.9848	0.9866	0.9770	0.7551	0.7899	**0.9904**	0.9766	0.9894
051938	**111812**	0.8898	0.8921	0.9272	0.8696	0.5611	0.6180	0.9284	0.9265	**0.9899**
051938	**120208**	0.9316	0.9578	0.9639	0.9490	0.7297	0.7839	0.9816	0.9729	**1.0000**
051938	**120902**	0.9500	0.9628	0.9685	0.9477	0.6746	0.7498	0.9587	0.9780	**0.9959**
051938	**121215**	0.8662	0.8758	0.8012	0.8156	0.6254	0.6887	0.8657	0.9208	**0.9369**
071450	**081358**	0.9721	0.9825	0.9840	0.9787	0.5578	0.6209	0.9628	0.9803	**0.9912**
071450	**111812**	0.8774	0.8862	0.9275	0.8101	0.5214	0.6305	0.9502	0.9119	**0.9929**
071450	**120208**	0.9321	0.9617	0.9635	0.9371	0.5482	0.5912	0.9332	0.9699	**0.9869**
071450	**120902**	0.9398	0.9480	0.9408	0.9130	0.5624	0.6599	0.9294	0.9628	**0.9863**
071450	**121215**	0.8363	0.8587	0.7582	0.7194	0.5077	0.5491	0.8065	0.8905	**0.9079**
081358	**111812**	0.8842	0.8854	0.9189	0.8306	0.5548	0.6565	0.9293	0.9021	**0.9955**
081358	**120208**	0.9290	0.9336	0.9467	0.9311	0.7309	0.8108	0.9633	0.9649	**0.9794**
081358	**120902**	0.9479	0.9587	0.9369	0.9326	0.7041	0.7778	0.9517	0.9558	**0.9826**
081358	**121215**	0.8504	0.8798	0.7718	0.7618	0.6688	0.7585	0.8669	0.9047	**0.9270**
111812	**120208**	0.8854	0.9102	0.9662	0.9043	0.6948	0.8060	0.9580	0.9257	**0.9862**
111812	**120902**	0.8764	0.8995	0.9441	0.8638	0.7611	0.8371	0.9336	0.9067	**0.9887**
111812	**121215**	0.8018	0.8212	0.8022	0.7367	0.7029	0.8166	0.8748	0.8023	**0.9141**
120208	**120902**	0.9901	0.9942	0.9962	0.9909	0.7547	0.8292	0.9954	0.9963	**1.0000**
120208	**121215**	0.9251	0.9402	0.8573	0.9156	0.7642	0.8697	0.9051	**0.9475**	0.9466
120902	**121215**	0.9316	0.9415	0.8486	0.9225	0.7861	0.8535	0.9023	0.9436	**0.9449**
Average recall@1		**0.8851**	**0.8999**	**0.8927**	**0.8314**	**0.6447**	**0.7305**	**0.9389**	**0.9421**	**0.9796**

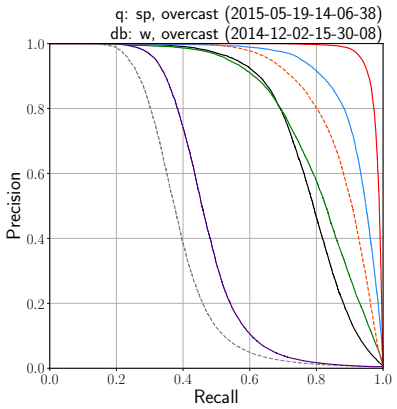

q: w, dawn, snow (2015-02-03-08-45-10)
db: w, dawn (2015-02-13-09-16-26)
q: w, dawn, snow (2015-02-03-08-45-10)
db: sp, sun, roadwork (2015-03-10-14-18-10)
q: w, dawn, snow (2015-02-03-08-45-10)
db: sp, overcast (2015-05-19-14-06-38)
q: w, dawn, snow (2015-02-03-08-45-10)
db: su, sun (2014-07-14-14-49-50)
q: w, dawn, snow (2015-02-03-08-45-10)
db: su, overcast (2015-08-13-16-02-58)
q: w, dawn, snow (2015-02-03-08-45-10)
db: a, sun (2014-11-18-13-20-12)
q: w, dawn, snow (2015-02-03-08-45-10)
db: w, overcast (2014-12-02-15-30-08)
q: w, dawn, snow (2015-02-03-08-45-10)
db: w, overcast (2014-12-09-13-21-02)
q: w, dawn, snow (2015-02-03-08-45-10)
db: w, overcast (2014-12-12-10-45-15)
q: w, dawn (2015-02-13-09-16-26)
db: sp, sun, roadwork (2015-03-10-14-18-10)
q: w, dawn (2015-02-13-09-16-26)
db: sp, overcast (2015-05-19-14-06-38)
q: w, dawn (2015-02-13-09-16-26)
db: su, sun (2014-07-14-14-49-50)
NetVLAD (ots)
NetVLAD (ft)
SeqSLAM (20 m)
SeqSLAM (40 m)
Appearance descr.
Structure descr.
Composite descr. (ours)
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

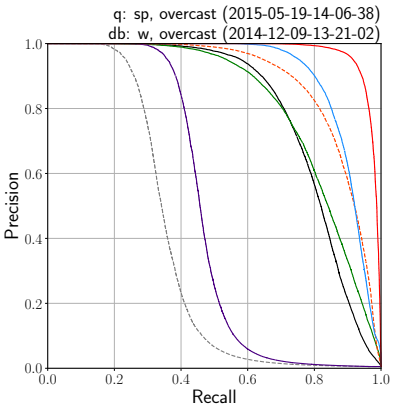

q: w, dawn (2015-02-13-09-16-26)
db: su, overcast (2015-08-13-16-02-58)
q: w, dawn (2015-02-13-09-16-26)
db: a, sun (2014-11-18-13-20-12)
q: w, dawn (2015-02-13-09-16-26)
db: w, overcast (2014-12-02-15-30-08)
q: w, dawn (2015-02-13-09-16-26)
db: w, overcast (2014-12-09-13-21-02)
q: w, dawn (2015-02-13-09-16-26)
db: w, overcast (2014-12-12-10-45-15)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: sp, overcast (2015-05-19-14-06-38)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: su, sun (2014-07-14-14-49-50)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: su, overcast (2015-08-13-16-02-58)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: a, sun (2014-11-18-13-20-12)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: w, overcast (2014-12-02-15-30-08)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: w, overcast (2014-12-09-13-21-02)
q: sp, sun, roadwork (2015-03-10-14-18-10)
db: w, overcast (2014-12-12-10-45-15)
NetVLAD (ots)
NetVLAD (ft)
SeqSLAM (20 m)
SeqSLAM (40 m)
Appearance descr.
Structure descr.
Composite descr. (ours)
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

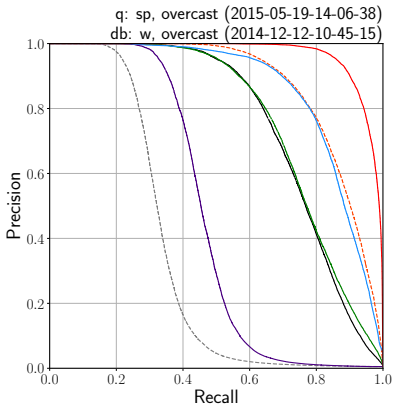

q: sp, overcast (2015-05-19-14-06-38)
db: su, sun (2014-07-14-14-49-50)
NetVLAD (ots)
NetVLAD (ft)
SeqSLAM (20 m)
SeqSLAM (40 m)
Appearance descr.
Structure descr.
Composite descr. (ours)
Precision
Recall

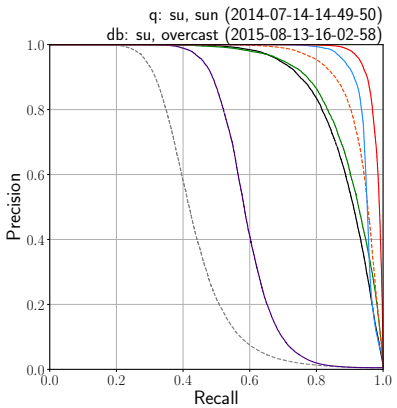

q: sp, overcast (2015-05-19-14-06-38)
db: su, overcast (2015-08-13-16-02-58)
Precision
Recall

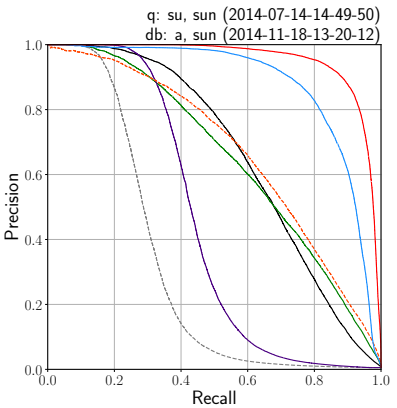

q: sp, overcast (2015-05-19-14-06-38)
db: a, sun (2014-11-18-13-20-12)
Precision
Recall

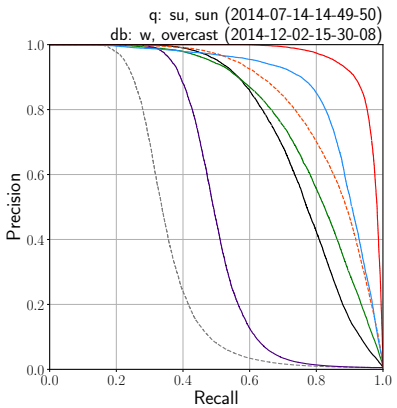

q: sp, overcast (2015-05-19-14-06-38)
db: w, overcast (2014-12-02-15-30-08)
Precision
Recall

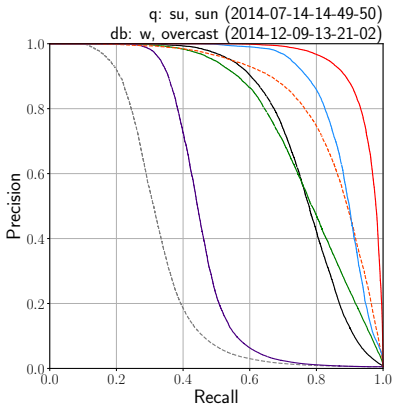

q: sp, overcast (2015-05-19-14-06-38)
db: w, overcast (2014-12-09-13-21-02)
Precision
Recall

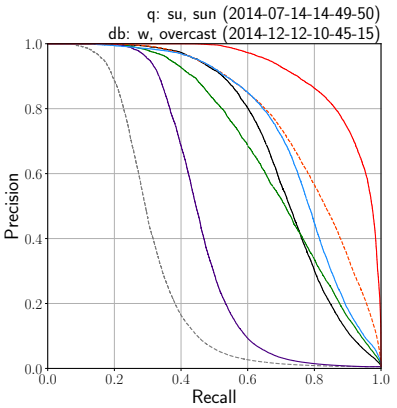

q: sp, overcast (2015-05-19-14-06-38)
db: w, overcast (2014-12-12-10-45-15)
Precision
Recall

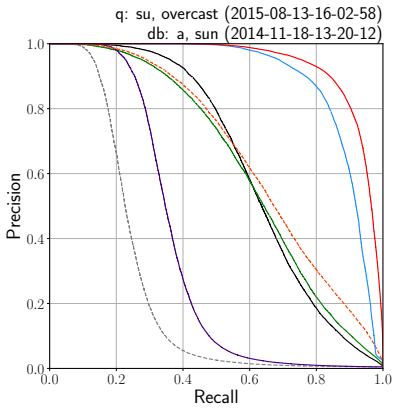

q: su, sun (2014-07-14-14-49-50)
db: su, overcast (2015-08-13-16-02-58)
Precision
Recall


q: su, sun (2014-07-14-14-49-50)
db: a, sun (2014-11-18-13-20-12)
Precision
Recall


q: su, sun (2014-07-14-14-49-50)
db: w, overcast (2014-12-02-15-30-08)
Precision
Recall


q: su, sun (2014-07-14-14-49-50)
db: w, overcast (2014-12-09-13-21-02)
Precision
Recall


q: su, sun (2014-07-14-14-49-50)
db: w, overcast (2014-12-12-10-45-15)
Precision
Recall


q: su, overcast (2015-08-13-16-02-58)
db: a, sun (2014-11-18-13-20-12)
Precision
Recall

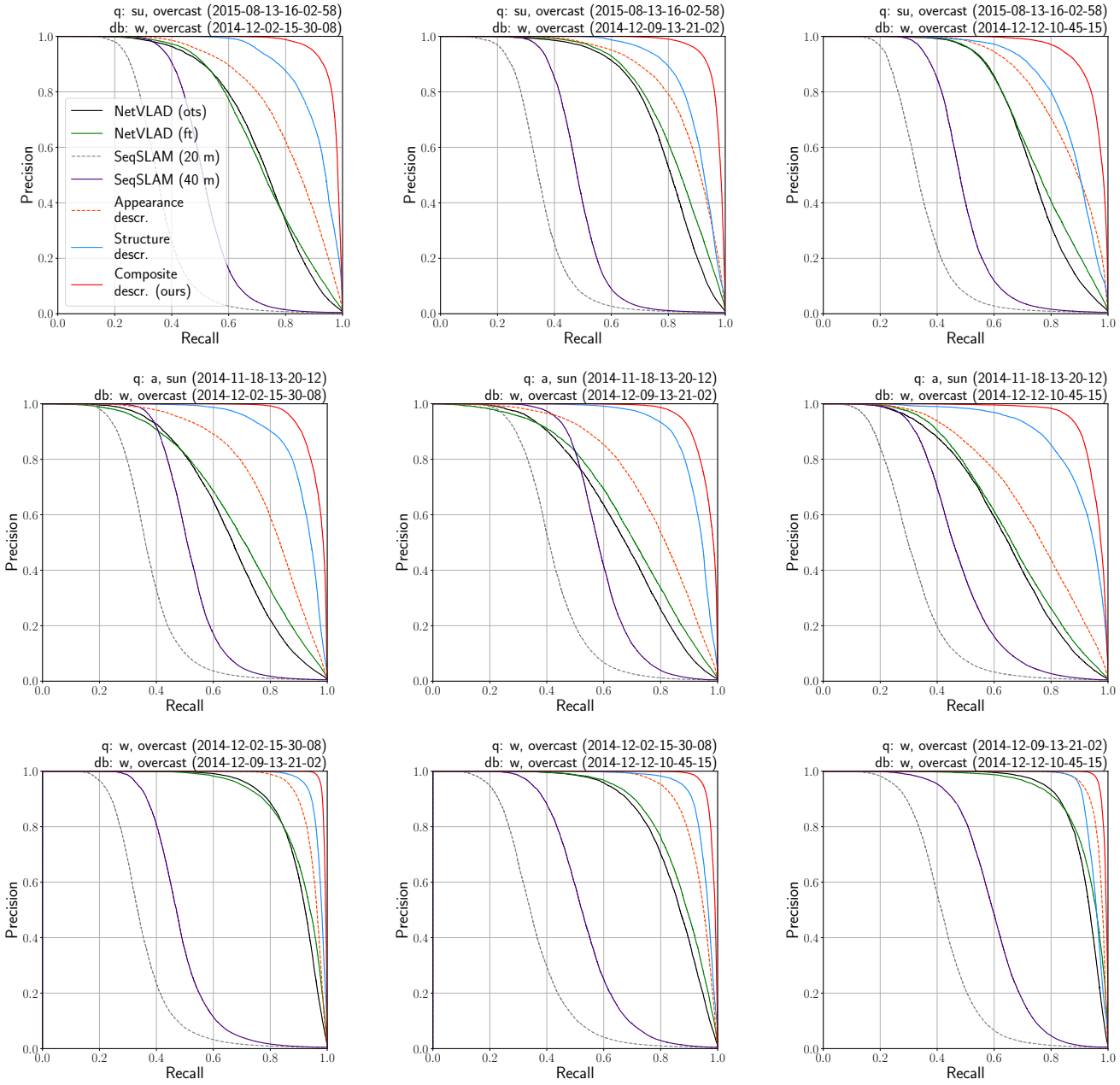


Fig. 5. Precision-recall curves for each pairing of query and database sequences out of a total of 45 sequence combinations. Each curve shows precision and recall, parametrized on the distance threshold $d_{emb,th}$ in embedding space, for exhaustive pairwise matching of query to database sequences. Each plot further details the varying conditions for both query and database sequences which are referenced using their timestamps. We compare our composite descriptors – obtained using a simple concatenation for combining intermediary visual and structural descriptors – to descriptors using only one of the input modalities as well as to off-the-shelf (ots) and fine-tuned (ft) variants of NetVLAD descriptors, as detailed in Section V-B. Specifically, pairwise matching based on our composite descriptor outperforms all other descriptors for *every* sequence combination.

Fig. 6. Examples of top retrieved database instances using appearance, structure, or composite descriptors. Only images of the retrieved instances are shown. Green and red borders indicate correct and incorrect matches. While descriptors that separately leverage either appearance or structure often produce incorrect matches, each query is successfully matched using our composite descriptor despite changes in viewpoint and visual appearance.