Supplementary Material
Unsupervised Adaptive Re-identification in Open World Dynamic Camera Networks

Rameswar Panda[1,*] Amran Bhuiyan[2,*,†] Vittorio Murino[2] Amit K. Roy-Chowdhury[1]
[1] Department of ECE, UC Riverside
[2]Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Italy
{rpand002@,amitrc@ece.}ucr.edu {amran.bhuiyan,vittorio.murino}@iit.it

Table 1. TABLE OF CONTENTS

Section No.	**Contents**
[2]	Detailed Information on the Experimented Datasets
[3]	Re-identification by Introducing a New Camera: Individual CMC Curves
[5]	Re-identification by Introducing Multiple Cameras: Analysis with Multiple Best Source Cameras
[6]	Effect of Feature Representation: Analysis with WHOS Feature
[7]	Effect of Subspace Dimension in Computing Geodesic Flow Kernels
[8]	Qualitative Matching Results

*RP and AB should be considered as joint first authors
†This work was done while AB was a visiting student at UC Riverside

Detailed Information on the Experimented Datasets

(a) WARD

(b) RAiD

(c) SAIVT-SoftBio

(e) Shinpuhkan2014

Figure 1. 24 image pairs from the (a) WARD, (b) RAiD, (c) SAIVT-SoftBio and (d) Shinpuhkan2014 datasets. Columns correspond to different persons, rows to different cameras. As seen from the figures, person re-identification is a challenging problem due to viewpoint changes, occlusions, illumination changes and background clutter in images of the same person in cameras with non-overlapping fields of view. Best viewed in color.

In this section, we describe the details on the four experimented datasets, namely WARD[1], RAiD[2], SAIVT-SoftBio[3] and Shinpuhkan2014[4] (See Fig. 1 for few sample images). As explained in the main paper, although there are number of other datasets (*e.g.* ViPeR, CAVIAR4REID, PRID450S, and CUHK etc.) for evaluating the performance in re-id, these datasets do not fit our purposes since they have only two cameras or specifically designed for video-based re-identification.

WARD [6] has 4786 images of 70 different people captured in a real surveillance scenario from 3 non-overlapping cameras. This dataset has a huge illumination variation apart from resolution and pose changes.

RAiD [3] was collected with a view to have large illumination variation that is not present in most of the publicly available benchmark datasets. In the original dataset 43 subjects were asked to walk through 4 cameras of which two are outdoor and two are indoor to make sure there is enough variation of appearance between cameras.

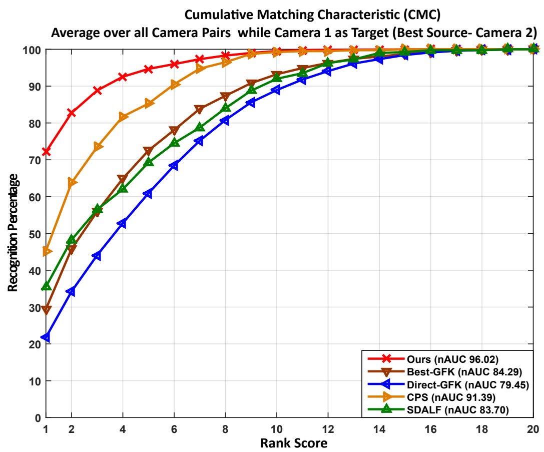
SAIVT-SoftBio [1] includes annotated sequences (704×576 pixels, 25 frames per second) of 150 people, each of which is captured by a subset of 8 different cameras placed inside an institute, providing various viewing angles and varying illumination conditions.

Shinpuhkan2014 [4] dataset consists of more than 22,000 images of 24 people which are captured by 16 cameras installed in a shopping mall. All images are manually cropped and resized to 48×128 pixels, grouped into tracklets with annotation. The number of tracklets of each person is 86. This dataset contains multiple tracklets in different directions for each person within a camera. To the best of our knowledge, this is the largest publicly available dataset for re-id with 16 cameras.

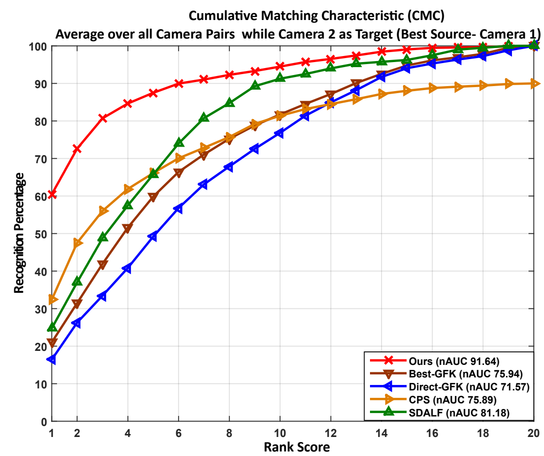
[1] http://www.ee.ucr.edu/~adas/raid.php
[2] http://users.dimi.uniud.it/~niki.martinel/code-and-datasets/
[3] https://esoe.qut.edu.au/qut-login/
[4] http://www.mm.media.kyoto-u.ac.jp/en/datasets/shinpuhkan/

Re-identification by Introducing a New Camera: Individual CMC Curves

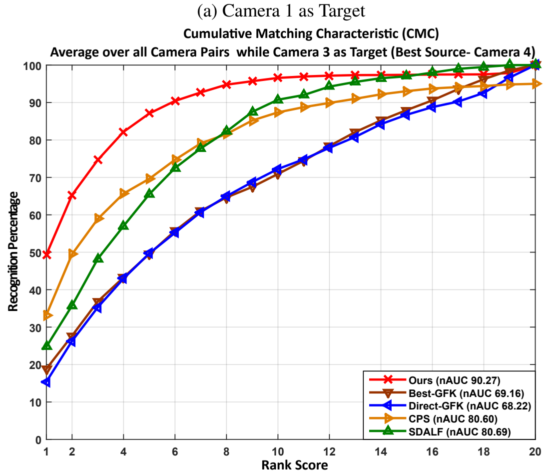
Individual CMC Curves for RAiD Dataset



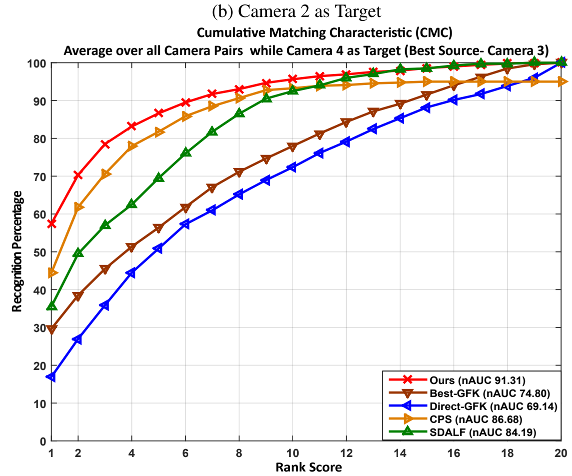
(a) Camera 1 as Target



(b) Camera 2 as Target



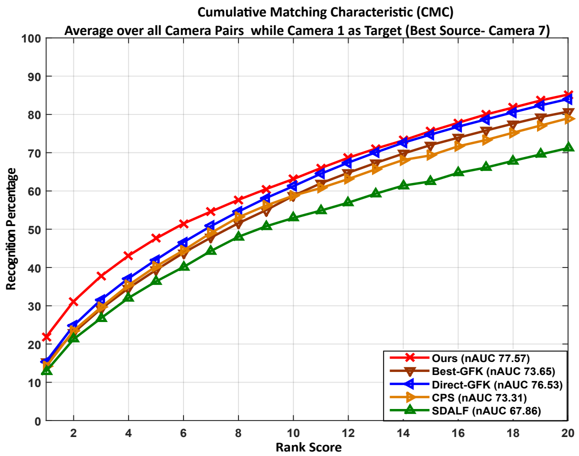
(c) Camera 3 as Target



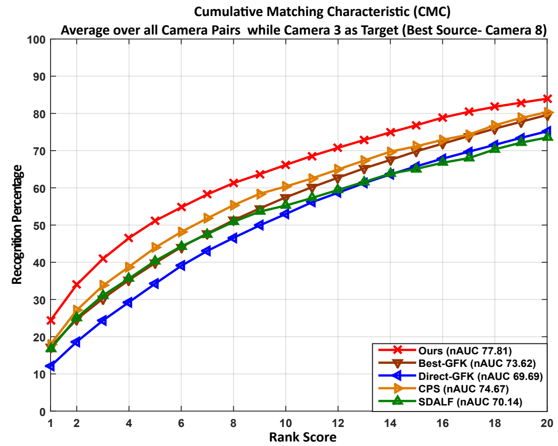
(e) Camera 4 as Target

Figure 2. CMC curves for RAiD dataset with 4 cameras. Plots (a, b, c, d) show the performance of different methods while introducing camera 1, 2, 3 and 4 respectively to a dynamic network. The proposed method significantly outperforms all the compared baselines for each case of the dynamic network. `CPS` is the closest baseline and works better than domain adaptive methods (`Best-GFk`, `Direct-GFk`) for most of the cases. `SDALF` is also competitive to `CPS` and outperforms in 2 cases (a, b) where camera 2 and 3 are introduced respectively. The performance gap between `Ours` and `Best-GFK` shows that the proposed transitive algorithm is effective in exploiting information from the best source camera while computing re-id accuracy across camera pairs. Best viewed in color.

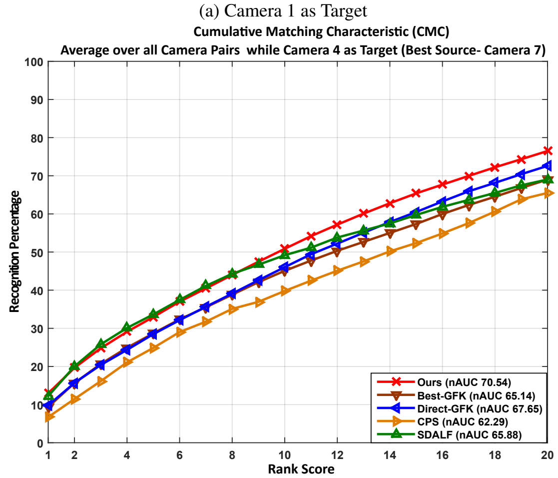
Individual CMC Curves for SAIVT-SoftBio Dataset



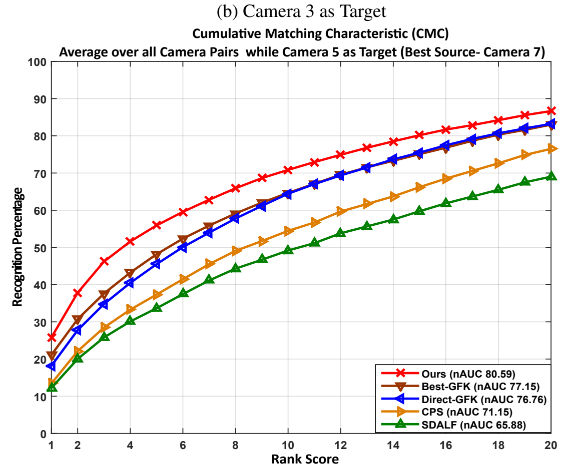
(a) Camera 1 as Target



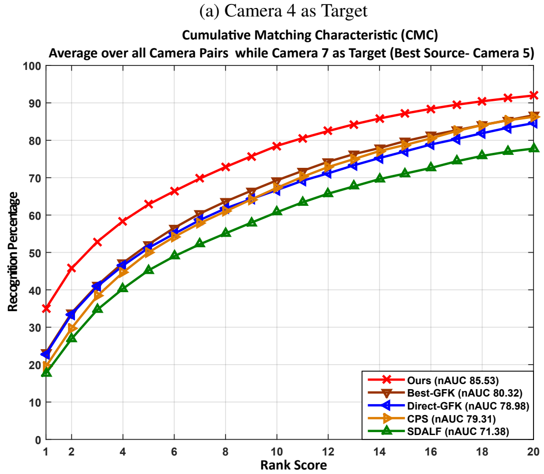
(b) Camera 3 as Target



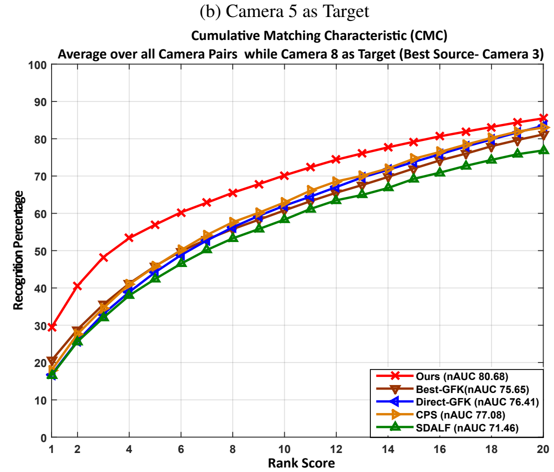
(a) Camera 4 as Target



(b) Camera 5 as Target



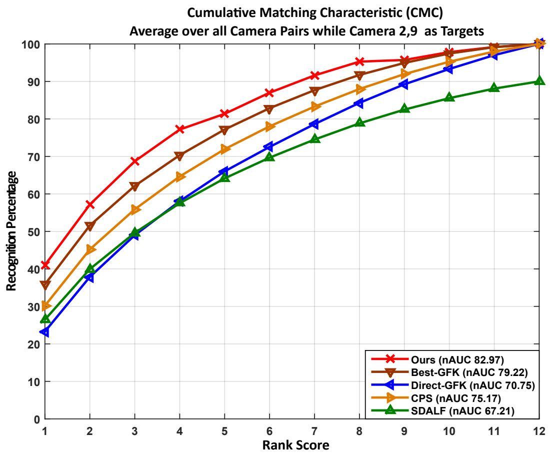
(a) Camera 7 as Target



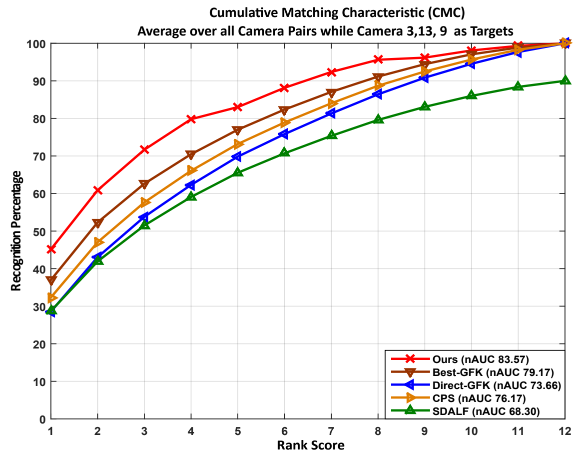
(b) Camera 8 as Target

Figure 3. CMC curves for SAIVT-SoftBio dataset with 8 cameras. Plots (a, b, c, d, e, f) show the performance of different methods while introducing camera 1, 3, 4, 5, 7 and 8 respectively to a dynamic network. Since not every person appears in all cameras, following the standard evaluation setting in [7, 2], we select those appearing in six cameras (1, 3, 4, 5, 7 and 8) as our evaluation set. The proposed method significantly outperforms all the compared baselines for each case of the dynamic network. Best viewed in color.

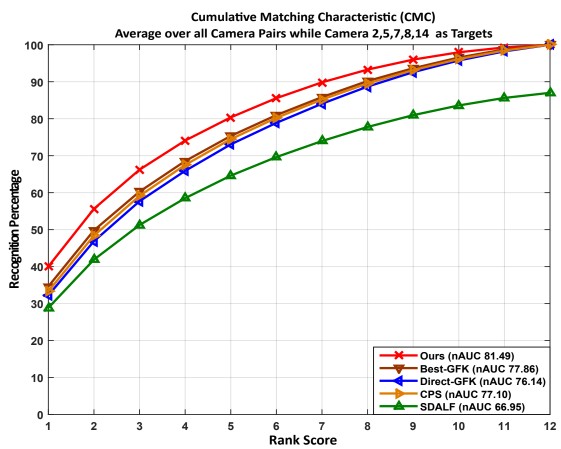
Re-identification by Introducing Multiple Cameras: Analysis with Multiple Best Source Cameras



(a) Camera 2, 9 as Targets



(b) Camera 3, 9, 13 as Targets



(c) Camera 2, 5, 7, 8, 14 as Targets

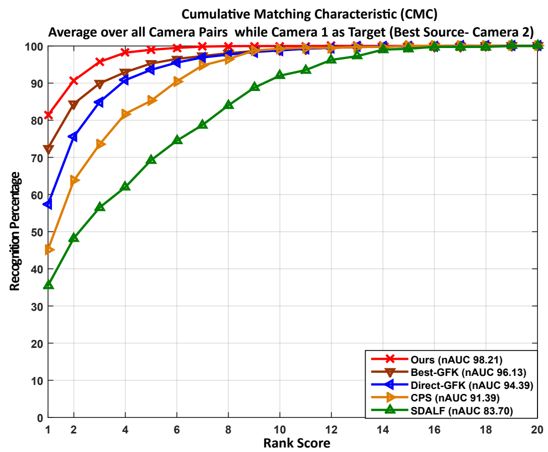
Figure 4. CMC curves for Shinpuhkan2014 dataset with 16 cameras. Plots (a, b, c) show the performance of different methods while introducing 2, 3 and 5 cameras respectively at the same time. We use multiple best source cameras, one for each target camera while computing re-id performance across a network. Please see the text for the analysis of the results.

Goal. The objective of this experiment is to validate the effectiveness of our approach while introducing multiple cameras at the same time in a dynamic camera network. Specifically, our goal is to show the performance of our method with multiple best source cameras, one for each target camera instead of a common best source camera as in Section 4.3 of the main paper.

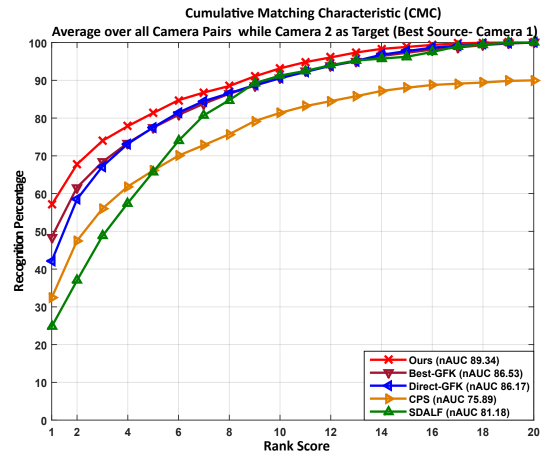
Implementation Details. We conduct this experiment on Shinpuhkan2014 dataset with of 16 cameras. Similar to the experiment in Section 4.3 of the main paper, we chose same 2, 3 and 5 cameras as the target cameras while remaining cameras as possible source cameras. However, instead of using the common best source camera, we use multiple best source cameras, one for each target camera in the transitive inference.

Results. We have the following key findings from Fig. 4: (i) Similar to the results in Section 4.3 of the main paper, our approach still outperforms all compared methods in all three scenarios. This indicates that the proposed method is very effective and can be applied to large-scale dynamic camera networks where multiple cameras can be introduced at the same time. (ii) The proposed adaptation approach works slightly better with multiple best source cameras compared to a common best source camera used for transitive inference. This is expected since multiple best source cameras can better exploit information from best source camera to improve the re-id accuracy. (iii) Our approach is generic which can handle either multiple best source cameras or a common best source camera for transitive inference in a dynamic camera network.

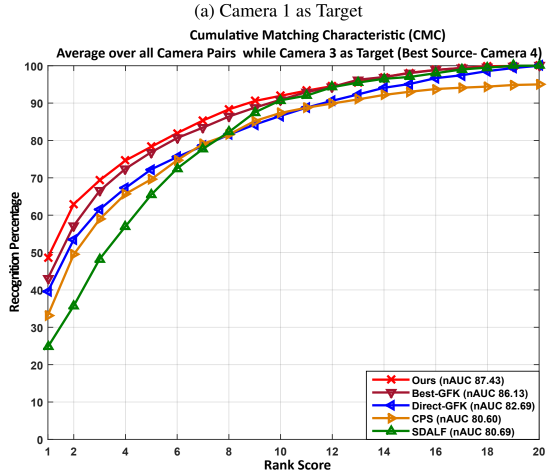
Effect of Feature Representation: Analysis with WHOS Feature



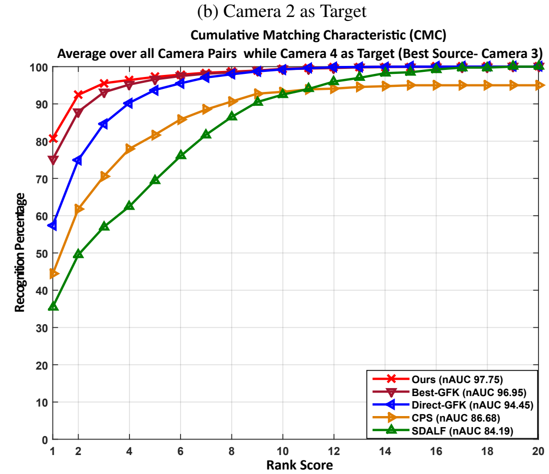
(a) Camera 1 as Target



(b) Camera 2 as Target



(c) Camera 4 as Target



(e) Camera 4 as Target

Figure 5. Re-id performance with WHOS feature representation. Plots (a,b,c,d) show CMC curves averaged over all camera pairs while introducing camera 1, 2, 3 and 4 respectively to a dynamic network. Please see the text for analysis of the results. Best viewed in color.

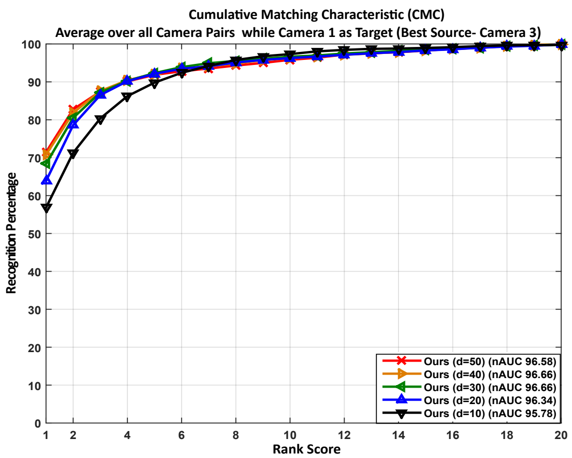
Goal. The objective of this experiment is to verify the effectiveness of our approach by changing the feature representation. Specifically, our goal is to show the performance of the proposed method by replacing LOMO feature with Weighted Histograms of Overlapping Stripes (WHOS) feature representation [5]. Ideally, we would expect similar performance improvement by our method, irrespective of the feature used to represent each person in a network of cameras.

Implementation Details. We use the publicly available code of WHOS[5] to test the performances and set the parameters same as recommended in the published work. Except the change in feature, we followed the same experimental settings while comparing with other methods.

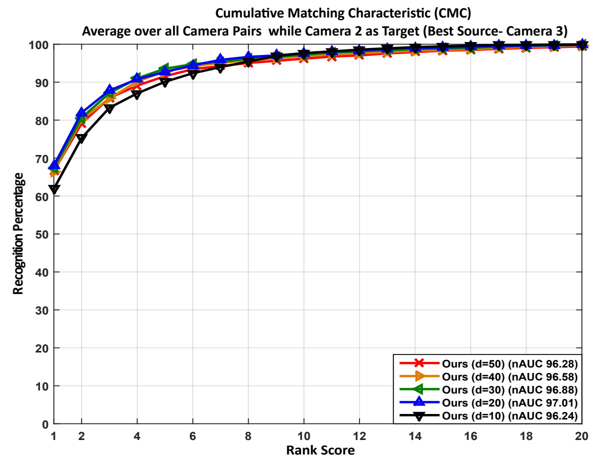
Results. Fig. 5 shows results for all possible 4 combinations (three source and one target) on RAiD dataset. From Fig. 5, the following observations can be made: (i) our approach outperforms all compared methods which suggests that the proposed adaptation technique works significantly well irrespective of the feature used to represent persons in a camera network. (ii) Among the alternatives, `Best-GFK` is the most competitive. However, the gap is still significant compared to `Ours` with an average margin of about 10%. (iii) The improvement over `Best-GFK` shows that the proposed transitive algorithm is very effective in exploiting information from the best source camera irrespective of the feature representation.

[5] http://www.micc.unifi.it/lisanti/source-code/whos/

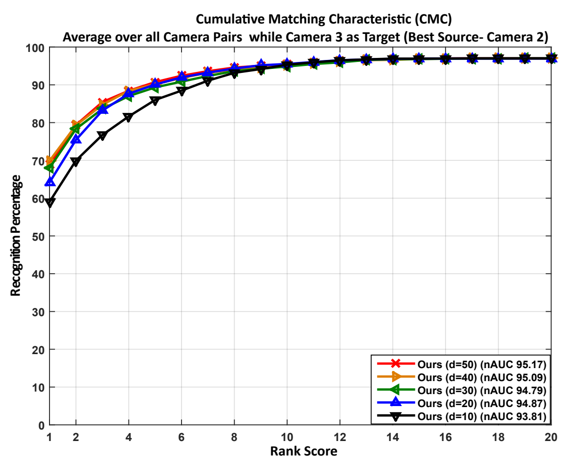
<u>**Effect of Subspace Dimension in Computing Geodesic Flow Kernels**</u>



(a) Camera 1 as Target



(b) Camera 2 as Target



(c) Camera 3 as Target

Figure 6. CMC curves for WARD dataset with 3 cameras. Plots (a, b, c) show the performance of different methods while introducing camera 1, 2 and 3 respectively to a dynamic network. Please see the text for the analysis of the results. Best viewed in color.

Goal. The main objective of this experiment is to analyze the performance of our method by changing the dimension of subspace used to compute the geodesic flow kernels across target and source cameras. In ideal case, we would like to see a minor change in performance with changing the dimension of subspace.

Implementation Details. We tested our approach with 5 cases of d, set to 10, 20, 30, 40 and 50. Except the change in dimension, we kept everything fixed while computing re-id performance in a dynamic camera network.

Results. We have the following key observations from Fig. 6: (i) Dimensionality of the subspace has a little effect on the re-id performance of our method. This suggests that the proposed method is robust to the change in dimensionality of the subspace used to compute the geodesic kernels across target and source cameras. (ii) Performance of our method is comparatively less when the dimension is set to 10. We believe this is because the principal angles computed at a dimension of 10 for this dataset are very small in magnitude which suggests that variances captured in the subspace corresponding to the source cameras would not be able to transfer to the target subspace. (iii) Although we empirically set the dimension to 50 in all our experiments, we believe finding the optimal dimension specific to a dataset can provide best re-id performance. We leave this problem of selecting optimal dimensionality of the subspace as future work.

Qualitative Matching Results

WARD Dataset

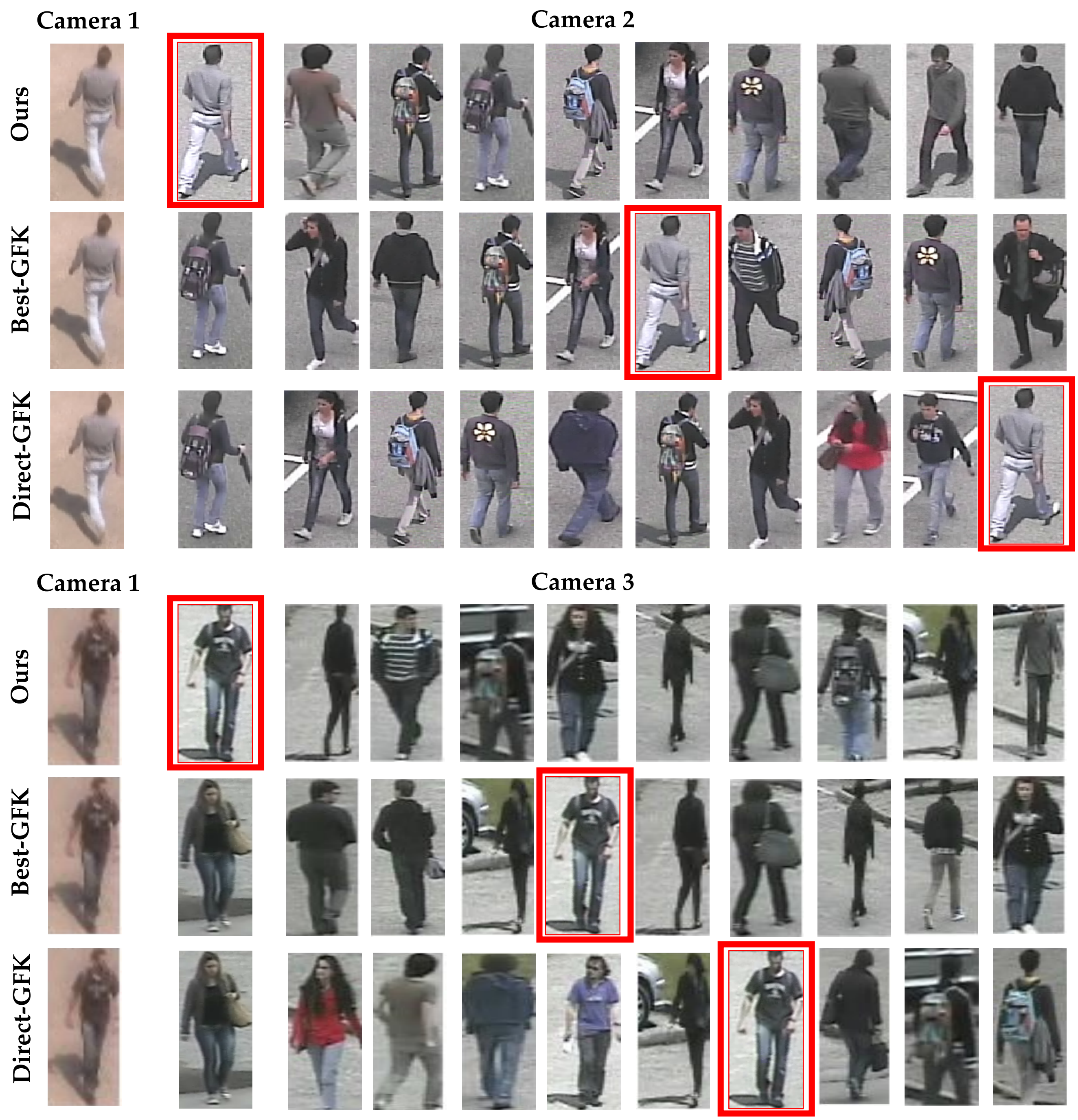


Figure 7. Visual comparison of two random persons from a newly introduced camera to top 10 matches from an already introduced camera in **WARD dataset**. **Top row:** Our matching result across camera pairs using the transitive algorithm. **Middle row:** matching the same person using `Best-GFK`. **Bottom row:** matching the same person using `Direct-GFK`. As can be seen from both figures, `Ours` perform best in matching persons (both in rank-1) across camera pairs by exploiting information from the best source camera. Best viewed in color.

Figure 8. Visual comparison of two random persons from a newly introduced camera to top 10 matches from an already introduced camera in **RAiD dataset**. **Top row:** Our matching result across camera pairs using the transitive algorithm. **Middle row:** matching the same person using `Best-GFK`. **Bottom row:** matching the same person using `Direct-GFK`. As can be seen from both figures, `Ours` perform best in matching persons (within rank-2) across camera pairs by exploiting information from the best source camera. Best viewed in color.

Figure 9. Visual comparison of two random persons from a newly introduced camera to top 10 matches from an already introduced camera in **SAIVT-SoftBio dataset**. **Top row:** Our matching result across camera pairs using the transitive algorithm. **Middle row:** matching the same person using `Best-GFK`. **Bottom row:** matching the same person using `Direct-GFK`. As seen from both figures, `Ours` perform best in matching persons (within rank-2) across camera pairs by exploiting information from the best source camera. Best viewed in color.

Figure 10. Visual comparison of two random persons from a newly introduced camera to top 10 matches from an already introduced camera in **Shinpuhkan2014 dataset**. **Top row:** Our matching result across camera pairs using the transitive algorithm. **Middle row:** matching the same person using `Best-GFK`. **Bottom row:** matching the same person using `Direct-GFK`. As seen from both figures, `Ours` perform best in matching persons (both in rank-2) across camera pairs by exploiting information from the best source camera. Best viewed in color.

References

[1] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey. A database for person re-identification in multi-camera surveillance networks. In *DICTA*, 2012. 2

[2] B. Cancela, T. M. Hospedales, and S. Gong. Open-world person re-identification by multi-label assignment inference. 2014. 4

[3] A. Das, A. Chakraborty, and A. K. Roy-Chowdhury. Consistent re-identification in a camera network. In *ECCV*, 2014. 2

[4] Y. Kawanishi, Y. Wu, M. Mukunoki, and M. Minoh. Shinpuhkan2014: A multi-camera pedestrian dataset for tracking people across multiple cameras. In *20th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2014. 2

[5] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *TPAMI*, 2015. 6

[6] N. Martinel and C. Micheloni. Re-identify people in wide area camera network. In *CVPRW*, 2012. 2

[7] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao. Multi-task learning with low rank attribute embedding for person re-identification. In *ICCV*, 2015. 4