Joint Intensity and Spatial Metric Learning for Robust Gait Recognition -Supplementary Material-

Yasushi Makihara[1], Atsuyuki Suzuki[1], Daigo Muramatsu[1], Xiang Li[2,1], Yasushi Yagi[1]
[1]Osaka University, [2]Nanjing University of Science and Technology
{makihara, a-suzuki, muramatsu, li, yagi}@am.sanken.osaka-u.ac.jp

Abstract

This supplementary material provides additional experimental results as well as some implementation aspects of the proposed method, which cannot be accommodated in the main paper due to page limitation.

1. More experimental results

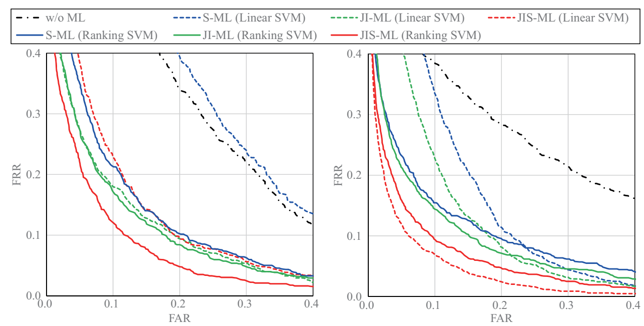
1.1. Analysis of individual modules

While we reported EER and rank-1 identification rate for analyzing individual modules, we provide ROC curves before and after z-normalization, and CMC curves both for OUTD-B and OU-LP-Bag β in Fig. 1 in this supplementary material. In addition, we also report summaries including not only EER and rank-1 identification rate but also area under curve (AUC) of the ROC curve and rank-5 identification rates of the CMC curve, both for training/testing sets and OUTD-B and OU-LP-Bag β in Table 1.

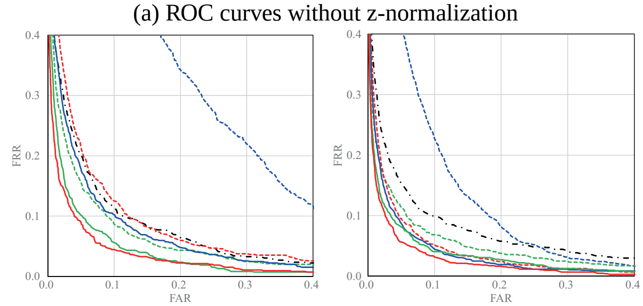
Similarly to the results reported in the main paper, it turns out that JIS-ML (Ranking SVM) yielded the best or the second best accuracies, and JIS-ML (Linear SVM) yielded the best accuracy for the test set of OU-LP-Bag β and for the training set in the verification mode (i.e., the lowest EER and AUC), which is reasonable by taking the properties of linear SVM and ranking SVM into account.

1.2. Sensitivity analysis

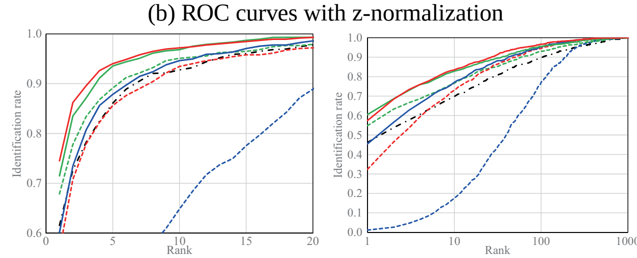
In this section, we provide results of sensitivity analysis of the hyper-parameters on the accuracies, i.e., EERs with z-normalization (denoted as z-EER). Specifically, we consider the soft margin parameter C in Eq. (16), the coefficient λ_S for regularizing the spatial metric in Eq. (17), the coefficient λ_I for regularizing the joint intensity metric in Eq. (20), and down-sampling rate (i.e., the number of spatial bins, and the number of intensity bins) as the hyper-parameters, and show EERs with z-normalization over the hyper-parameters in Fig. 2.



(a) ROC curves without z-normalization



(b) ROC curves with z-normalization



(c) CMC curves

Figure 1. ROC and CMC curves for individual metric learning and solvers (left: OUTD-B, right: OU-LP-Bag β).

As a result, the accuracies do not significantly change within a range from 0.1 to 10 as for the soft margin parameter C (Fig. 2(a)). On the other hand, the accuracies drop as the regularization coefficients λ_S and λ_I decrease, in particular for the spatial regularization coefficient λ_S and in case of small number of training subjects (i.e., OUTD-B) as shown in Fig. 2(b). This demonstrates the necessity of the

Table 1. Results for individual metric learning and solvers. Rank-1 and Rank-5 indicate rank-1 and rank-5 identification rates [%], respectively. Bold and Italic bold indicate the best and the second best accuracies.

(a-1) OUTD-B (Test set)

Method	EER	AUC	Rank-1	Rank-5
w/o ML	26.3	17.6	61.6	86.1
S-ML (linear SVM)	27.2	19.5	12.0	43.9
JI-ML (linear SVM)	13.9	6.0	67.9	89.1
JIS-ML (linear SVM)	14.5	7.2	56.4	85.6
S-ML (rank SVM)	14.6	7.2	60.2	87.9
JI-ML (rank SVM)	***13.4***	***5.8***	***71.5***	***93.6***
JIS-ML (rank SVM)	**11.0**	**4.1**	**74.5**	**94.0**

(a-2) OUTD-B (Training set)

Method	EER	AUC	Rank-1	Rank-5
w/o ML	27.2	19.3	72.3	92.5
S-ML (linear SVM)	7.0	2.1	84.5	99.5
JI-ML (linear SVM)	14.6	6.1	79.1	93.9
JIS-ML (linear SVM)	**1.5**	**0.2**	96.9	**100.0**
S-ML (rank SVM)	7.9	2.6	***97.7***	**100.0**
JI-ML (rank SVM)	11.3	4.4	91.3	99.8
JIS-ML (rank SVM)	***3.5***	***0.6***	**99.8**	**100.0**

(b-1) OU-LP-Bag β (Test set)

Method	EER	AUC	Rank-1	Rank-5
w/o ML	24.8	16.1	46.2	63.0
S-ML (linear SVM)	16.9	9.5	1.2	8.4
JI-ML (linear SVM)	12.1	4.9	54.8	70.6
JIS-ML (linear SVM)	**8.0**	**2.3**	32.4	63.1
S-ML (rank SVM)	13.3	6.0	45.5	69.1
JI-ML (rank SVM)	12.3	4.9	**60.5**	***76.9***
JIS-ML (rank SVM)	***9.8***	***3.3***	***57.4***	**77.7**

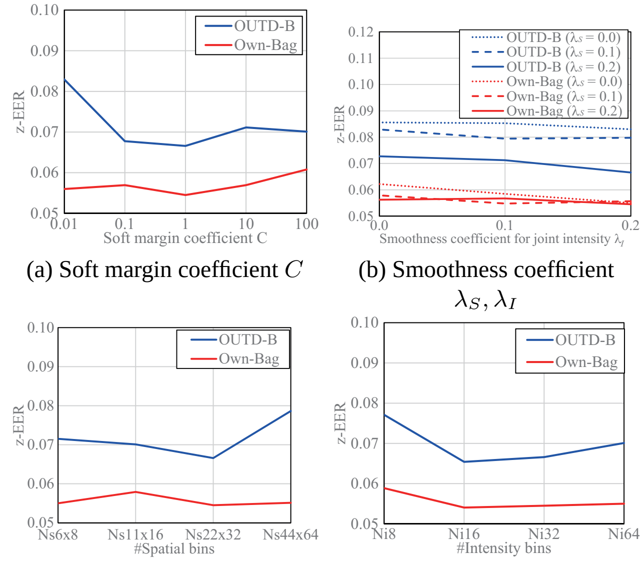
(b-2) OU-LP-Bag β (Training set)

Method	EER	AUC	Rank-1	Rank-5
w/o ML	23.9	16.0	48.3	64.5
S-ML (linear SVM)	14.0	7.4	1.7	9.3
JI-ML (linear SVM)	12.2	5.6	55.2	70.1
JIS-ML (linear SVM)	**4.8**	**1.1**	37.2	65.3
S-ML (rank SVM)	11.4	4.9	43.9	68.3
JI-ML (rank SVM)	12.4	5.2	**60.6**	***75.9***
JIS-ML (rank SVM)	***8.5***	***2.6***	***58.3***	**80.3**

regularization for the proximity to keep the generalization capability.

As for the number of spatial bins (Fig. 2(c)), we notice that 44×64 bins yielded the worst result for OUTD-B. This shows that use of too much spatial bins induces generalization errors for small number of training subjects, and hence relatively coarse spatial bins (e.g., less than 22×32 bins) are recommended.

As for the number of intensity bins (Fig. 2(d)), we can see that too small number of intensity bins (e.g., 8 bins) yielded the worst accuracies, while too much number of intensity bins (e.g., 64 bins) does not improve the accuracies. Therefore, moderate number of intensity bins such as 16 or 32 bins suffice based on the trade-off between the accuracy



(a) Soft margin coefficient C (b) Smoothness coefficient λ_S, λ_I

(c) #Spatial bins (d) #Intensity bins

Figure 2. Sensitivity analysis of the hyper-parameters on EERs with z-normalization.

and computational cost.

2. Implementation aspects

2.1. Fusion of multiple samples

When a probe and a gallery contain multiple samples (e.g., multiple periods of gait features for gait recognition), we can mitigate the effect of noisy observations (foreground segmentation errors, gait fluctuations among periods, etc.) by a statistical fusion scheme. For this purpose, we simply compute a dissimilarity for each combination of probe and gallery samples based on Eq. (8) in the original paper, and then take an average over the combinations as a fused dissimilarity score.

2.2. Image registration

Before extracting a gait feature such as GEI, we register a silhouette image sequence along with the horizontal axis so as that a horizontal gravity center of a silhouette region can coincide with the image center. Since the gravity center may vary depending on clothing and carrying objects, mis-alignment between a probe and a gallery may occur. We therefore introduce the image registration procedure for each matching pair as preprocess both for training and test phases. More specifically, we decide amount of horizontal shift of the probe so as to minimize l_1-norm between the gallery and the shifted probe.