Fast Person Re-identification via Cross-camera Semantic Binary Transformation

We are grateful for all respectful reviewers' suggestions, and are also encouraged by the positive comments. We will first respond to comments raised by all three reviewers, and then answer other questions for each reviewer, respectively.

To all reviewers' concern on the selection of non-hashing ReID methods for comparison in Table 2 and Table 3:

As all reviewers agreed, we didn't expect that the matching accuracy of our method (i.e., CSBT) can beat all non-hashing ReID approaches, since CSBT has its own advantage in the matching efficiency. Therefore, in Tables 2-3, our goal is to show that CSBT can significantly improve the matching accuracy of existing hashing based ReID models, whilst remarkably reducing the time and storage costs of non-hashing ones. As stated in lines 685-687, Metric Learning, Local Patches based Matching, Deep Learning, etc. are different kinds of non-hashing based matching strategies, which can cover almost all kinds of ReID approaches. Their representative methods are KISSME, MLF and DeepReID, respectively. By comparing to them, the effectiveness and efficiency of our method can be solidly validated. That's the reason why we chose to report the results of KISSME, MLF, DeepReID, etc. in Tables 2-3.

We have realized that Tables 2-3 might make some readers misunderstand that the matching accuracy of CSBT is higher than all non-hashing ReID models. Considering that there are so many publications on ReID, it would be too messy to list the results of all methods, which is also not allowed due to the space limitation. Thus, we chose to summarize them in lines 741-755. Of course, we do agree that the experiments of our submission can be more comprehensive by including more results of the state-of-the-art non-hashing methods in Tables 2-3. We will add the results of the non-hashing methods suggested by reviewers (as shown in Table A) to Tables 2-3 in the final version of our paper.

Table A. Performance of the proposed method compared to the state-of-the-art non-hashing based ReID approaches.

Method		VIPeR r=1	VIPeR Time (s)	VIPeR Mem. (KB)	CUHK03 (Manual) r=1	CUHK03 (Manual) Time (s)	CUHK03 (Manual) Mem. (KB)
Non Hashing	JSTL [1]	38.6	-	5.68e+03	75.3	-	5.18e+02
	GSCNN [2]	37.8	-	5.68e+03	-	-	-
	EDM [3]	40.9	-	5.68e+03	61.3	-	5.18e+02
	SCSP [4]	53.5	3.78e-02	2.96e+02	-	-	-
	NSL (LOMO) [41]*	42.3	-	6.66e+04	58.9	-	2.11e+04
	Best	**63.9** [9]*	-	1.20e+04	**80.2** [43]*	-	5.18e+02
	Ours	36.6	**1.68e-06**	**3.45e+01**	55.5	4.83e-07	**1.01e+01**

Method		CUHK01 (p=486) r=1	CUHK01 (p=486) Time (s)	CUHK01 (p=486) Mem. (KB)	CUHK01 (p=100) r=1	CUHK01 (p=100) Time (s)	CUHK01 (p=100) Mem. (KB)
Non Hashing	JSTL [1]	66.6	-	1.70e+04	-	-	-
	EDM [3]	-	-	-	86.6	-	3.48e+03
	NSL (LOMO) [41]*	65.0	-	2.05e+05	-	-	-
	Best	**76.5** [43]*	-	1.70e+04	**89.6** [43]*	-	3.48e+03
	Ours	51.2	9.57e-06	**9.85e+01**	74.3	2.26e-07	**4.69e+00**

Method		Market-1501 (Single Query) r=1	Market-1501 (Single Query) Time (s)	Market-1501 (Single Query) Mem. (KB)	CUHK03 (Detected) r=1	CUHK03 (Detected) Time (s)	CUHK03 (Detected) Mem. (KB)
Non Hashing	EDM [3]	-	-	-	52.1	-	5.18e+02
	SCSP [4]	51.9	1.21e+01	7.67e+04	-	-	-
	NSL (LOMO) [41]*	55.4	-	4.13e+06	53.7	-	2.11e+04
	Best	**65.9** [2]	-	4.32e+04	**68.1** [2]	-	5.18e+02
	Ours	42.9	**4.7e-04**	**1.52e+03**	49.7	4.83e-07	**1.01e+01**

(* means that the corresponding reference comes from our original submission.)

To all reviewers' concern on the parameter settings:

We fine-tuned all parameters by cross-validation on the training data. For β, γ, ν, we found that our method could yield the promising performance by using $\beta = 0.01$, $\gamma = 0.1$ and $\nu = 1 \times 10^{-5}$, which were adopted in all experiments. In Table 2, for the bit length L, we used 896, 832, 832 and 640 bits on VIPeR, CUHK01, CUHK03 and Market-1501, respectively. In terms of μ, we set it as the averaged squared Euclidean distances, i.e., $\mu = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, to compensate for the bias. We will make the parameter settings clearer in the final version.

To Reviewer 1's and Reviewer 2's suggestions on improving the hyphenations and presentation clarity:

Thanks a lot for your advices. We agree that your suggestions can help make the presentation clearer. We will carefully revise our submission accordingly.

To Reviewer 1's and Reviewer 2's concern on the statement in lines 320-321:

We intend to claim that "It is not guaranteed to achieve a globally optimal solution", rather than "There is no global optimal solution". We will revise it in the final version.

To Reviewer 1's concern on comparing the embedded features with CSBT:

The rank 1 accuracies of the embedded features on VIPeR, CUHK01 (p=486), CUHK03 and Market-1501 are 39.7%, 60.2%, 57.2% and 44.1%, respectively. In fact, both the embedding and the accurate hashing significantly affect the matching accuracy of hashing based ReID models.

To Reviewer 2's concern on the difference between CBI and the proposed CSBT:

CBI is a hashing method for ReID that only addresses the matching efficiency, while CSBT can further mitigate cross-caremera variations in raw data to improve the matching accuracy, by employing a subspace projection. The formulations of CBI and CSBT w.r.t. hashing are also different. We will make this point clearer in the final version.

To Reviewer 4's concern on the performance of our method on CUHK03 using detected images:

In Table A, we briefly report the performance of our method on CUHK03 by using detected images. It can be observed that the performance of CSBT slightly drops, due to the detection error. We will add more results on CUHK03 in the final version of our submission.

References

[1] Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*, 2016. 1

[2] Gated siamese convolutional neural network architecture for human re-identification. In *ECCV*, 2016. 1

[3] Embedding deep metric for person re-identification: A study against large variations. In *ECCV*, 2016. 1

[4] Similarity learning with spatial constraints for person re-identification. In *CVPR*, 2016. 1