Supplementary Material for
Hierarchical Gaussian Descriptor for Person Re-Identification

Tetsu Matsukawa[1], Takahiro Okabe[2], Einoshin Suzuki[1], Yoichi Sato[3]
[1] Kyushu University [2] Kyushu Institute of Technology [3] The University of Tokyo
{matsukawa, suzuki}@kyushu-u.ac.jp, okabe@ai.kyutech.ac.jp, ysato@iis.u-tokyo.ac.jp

1. Details of the baseline descriptors

In section 4.2 of the paper, we compared the distribution modeling of GOG to other distributions. Below, we describe the details of the compared methods.

The Mean, Cov and Gauss are global distribution descriptors of pixel features within each region. The Cov-of-Cov, Cov-of-Gauss and GOG are hierarchical distribution descriptors. The Cov-of-Cov uses covariance matrix in both patch and region modeling. The Cov-of-Gauss uses Gaussian for patch modeling and covariance matrix for region modeling.

For a fair comparison to GOG which is incorporated with patch weights, we adopted the weighted pooling for all descriptors. Formally,

Mean: $\boldsymbol{\mu}' = \frac{1}{\sum_{i \in \mathcal{G}} w_i} \sum_{i \in \mathcal{G}} w_i \boldsymbol{f}_i$,

Cov: $\boldsymbol{\Sigma}' = \frac{1}{\sum_{i \in \mathcal{G}} w_i} \sum_{i \in \mathcal{G}} w_i (\boldsymbol{f}_i - \boldsymbol{\mu}')(\boldsymbol{f}_i - \boldsymbol{\mu}')^T$,

Gauss: $\boldsymbol{P}' = |\boldsymbol{\Sigma}'|^{-\frac{1}{d+1}} \begin{bmatrix} \boldsymbol{\Sigma}' + \boldsymbol{\mu}'\boldsymbol{\mu}'^T & \boldsymbol{\mu}' \\ \boldsymbol{\mu}'^T & 1 \end{bmatrix}$,

where w_i is a weight of pixel i and determined in the same manner as w_s.

Cov-of-Cov: $\boldsymbol{\Xi} = \frac{1}{\sum_{s \in \mathcal{G}} w_s} \sum_{s \in \mathcal{G}} w_s (\boldsymbol{h}_s - \boldsymbol{\nu})(\boldsymbol{h}_s - \boldsymbol{\nu})^T$, where $\boldsymbol{h}_s = \mathrm{vec}(\log(\boldsymbol{\Sigma}_{\mathrm{s}}))$ and $\boldsymbol{\nu} = \frac{1}{\sum_{s \in \mathcal{G}} w_s} \sum_{s \in \mathcal{G}} w_s \boldsymbol{h}_s$.

Cov-of-Gauss: As per $\boldsymbol{\Sigma}^{\mathcal{G}}$ defined by Eq.(5).

The tangent space mapping using log-Euclidean and the half vectorization are commonly applied for all descriptors except Mean. The descriptors of regions are concatenated to form an image representation.

Table 1 (a) summarizes the dimensionality of each descriptor. We commonly used the same 7 regions as GOG and the fusion approach that concatenates meta descriptors extracted from 8 dimensional pixel features ($d = 8$) on RGB color space. Here, let us denote $D(d)$ as the dimension per region for a meta descriptor. Then the feature vector dimension of an image becomes 7 regions $\times D(8)$ dim. For example, the dimensionality of Cov becomes $7 \times (8^2 + 8)/2 = 252$. The dimensionality of other descriptors can be obtained with the same way.

Table 1. Dimensions of each meta descriptor. (a) For comparison in Sec.4.2. (b) For comparison in Sec.4.3.

	Methods	Dimension per region $D(d)$
(a)	Mean	d
	Cov	$m' = (d^2 + d)/2$
	Gauss	$m = (d^2 + 3d)/2 + 1$
	Cov-of-Cov	$(m'^2 + m')/2$
	Cov-of-Gauss	$(m^2 + m)/2$
	GOG	$(m^2 + 3m)/2 + 1$
(b)	GOLD	$d + (d^2 + d)/2$
	2AvgP	$(d^2 + d)/2$
	HASC	$d^2 + d$
	LDFV	$2Kd$

2. Details of other meta descriptors

In Table 1 of the paper, we compared the GOG descriptor with other meta descriptors. Below, we describe the details of these descriptors used in the comparison.

Cov [17]**:** The covariance descriptor describes an image region by a $d \times d$ dimensional covariance matrix. We applied log-Euclidean and half-vectorization to the covariance matrix. Therefore, the dimensionality is $(d^2 + d)/2$ per region.

HASC [2]**:** The HASC is composed of a covariance descriptor and an Entropy and Mutual Information (EMI) descriptor. The EMI descriptor captures the non-linear relation within pixel features and its dimensionality is the same as the covariance descriptor. Therefore, the dimensionality of the HASC is $d^2 + d$ per region. For implementation, we used the code provided by the authors [2].

GOLD [16]**:** The GOLD describes an image region by a mean vector and a covariance matrix. The covariance matrix is flattened by log-Euclidean and the half-vectorization is applied. The two components are concatenated into one vector. Therefore, the dimensionality is $d + (d^2 + d)/2$ per region. We omitted the spatial pyramid and the power normalization used in the original paper.

2AVgP [3] The 2AvgP describes an image region by a zero-mean covariance matrix, *i.e.*, autocorrelation matrix. We applied log-Euclidean and the half-vectorization to the matrix. Therefore, the dimensionality is $(d^2 + d)/2$ per region.

LDFV [10] **:** The LDFV encodes pixel features using Fisher Vector (FV) coding, which encodes the difference of pixel features from the pre-trained GMM means. The dimensionality is $2Kd$, where K is the number of GMM components. The parameters of GMM were estimated on each training set of random splits. Following the recommended setting [10], we set $K = 16$. For implementation of the GMM estimation and FV coding, we used VLFeat [18].

Cov-of-Cov [7, 15] **:** The Cov-of-Cov describes an image region as a covariance matrix of patch covariances. Each patch within a region is described as a $d \times d$ dimensional covariance matrix. The patch covariance matrix is flattened and half-vectorized into a $m' = (d^2 + d)/2$ dimensional vector. Each region is described as a $m' \times m'$ dimensional covariance matrix of the vectors of patch covariances. By applying log-Euclidean and the half-vectorization again, the dimensionality per region becomes $(m'^2 + m')/2$.

Table 1 (b) summarizes dimensionality of each meta descriptor. We commonly used the same 7 regions as GOG and the fusion approach that concatenates meta descriptors extracted from 4 pixel features, those dimensions are $d = \{8, 8, 8, 7\}$. Here, let us denote $D(d)$ as the dimension per region for a meta descriptor. Then the feature vector dimension of the image becomes 3 features $\times 7$ regions $\times D(8)$ dim. + 7 regions $\times D(7)$ dim. For example, the dimensionality of Cov becomes $3 \times 7 \times (8^2 + 8)/2 + 7 \times (7^2 + 7)/2 = 952$. The dimensionality of other descriptors can be obtained with the same way.

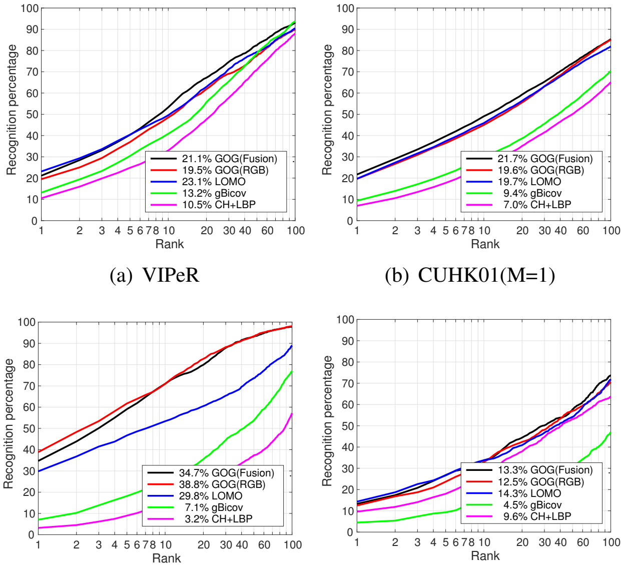
3. Performance without metric learning

To see the original performance of GOG without metric learning, we compare the performance with cosine distance. The compared descriptors are CH+LBP [20], gBiCov [11] and LOMO [8]. For all methods, the proposed mean removal + L2 normalization is adopted. Fig. 1 shows the CMC curves of the compared descriptors. It can be seen that the GOG$_{\text{Fusion}}$ achieves comparable results to LOMO in VIPeR, CHUK01 and GRID datasets. Besides, the GOG outperforms all compared descriptors on PRID450S dataset.

4. CMC curves for performance comparison

In Table 2 and Table 3 of the paper, we reported only the CMC scores at r = 1,5,10 and 20 due to the space limitation. In this supplementary material, we provide the full CMC curves over the rank 1 to 100 obtained by compared methods.

Fig 2 shows the CMC curves of the compared methods including GOG$_{\text{Fusion}}$+XQDA, MetricEmsemble [13], LOMO+XQDA [8], ImprovedDeep [1], SCNCD [21], SalMatch [22], MLFL [24], CH+LBP [20]+XQDA, and gBiCov [11]+XQDA . For several methods, the CMC curves are not included for several datasets because the previous works do not report them.



(a) VIPeR

(b) CUHK01(M=1)

(c) PRID450S

(d) GRID

Figure 1. CMC curves usnig cosine distance on (a) VIPeR, (b) CUHK01(M=1), (c) PRID450S and (d) GRID.

In addition to the compared methods in the main paper, the Figure contains methods including Local Fisher discriminant analysis (LF) [14], SDALF [4], KISSME [6], PCCA [12],PRDC [25], and ELF [5] on VIPeR dataset are brought from their original papers. The results of five ELF6 [5] features on GRID dataset are brought from the paper [9]. The results of SDALF [4], LMNN [19], eSDC [23], and KISSME [6] on CUHK01 and CUHK03 datasets are brought from the paper [1].

Acknowledgement

This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", under a fund for integrated promotion of social system reform and research and development of MEXT Japan, and JSPS KAKENHI 15K16028.

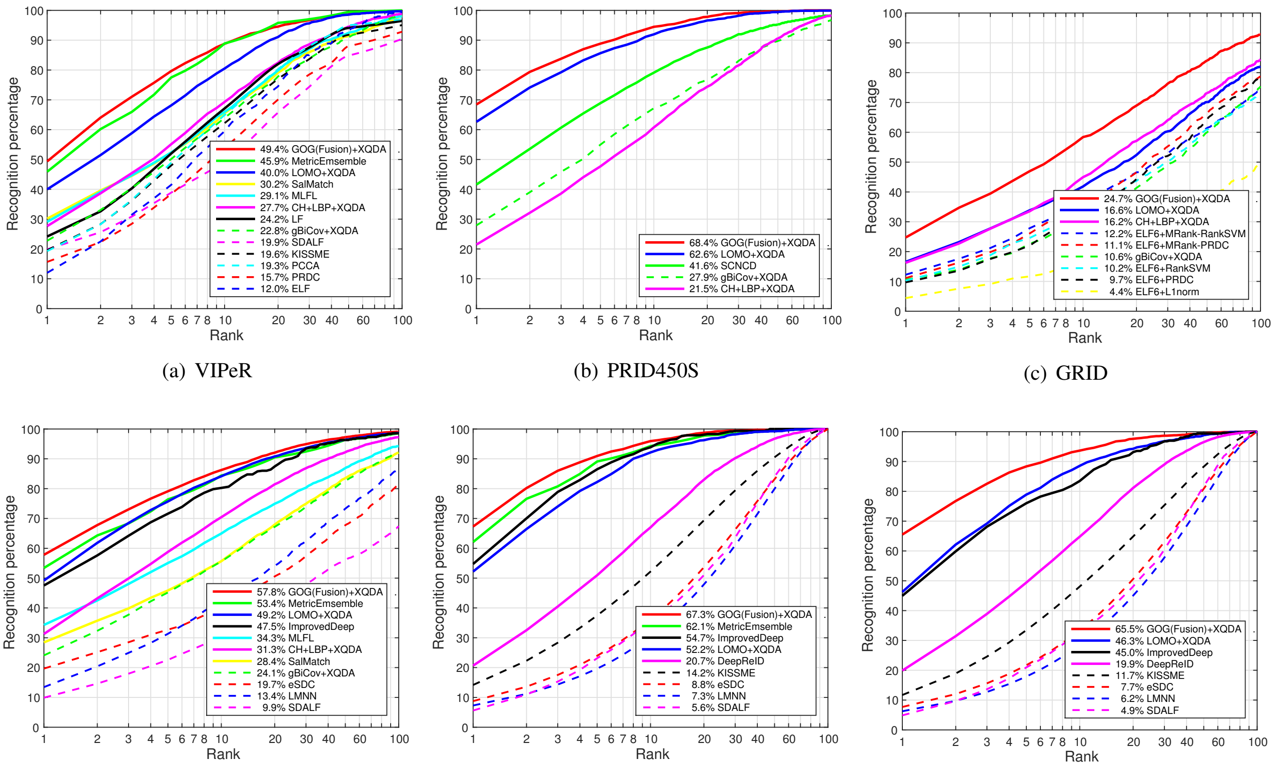
References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3908–3916, 2015. 2

[2] M. S. Biagio, M. Crocco, M. Cristani, S. Martelli, and V. Murino. Heterogeneous auto-similarities of characteristics (HASC): exploiting relational information for classification. In *IEEE International Conference on Computer Vision (ICCV)*, pages 809–816, 2013. 1

[3] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Free-form region description with second-order pooling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(6):1177–1189, 2015. 1

[4] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local



(a) VIPeR

(b) PRID450S

(c) GRID

(d) CUHK01(M=1)

(e) CUHK03 Labeled

(f) CUHK03 Detected

Figure 2. CMC curves of state-of-the-art methods on (a) VIPeR, (b) PRID450S, (c) GRID, (d) CHUK01(M=1), (e) CUHK03 Labeled and (f) CUHK03 Detected datasets.

features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2360–2367, 2010. 2

[5] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision (ECCV)*, pages 262–275, 2008. 2

[6] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2288–2295, 2012. 2

[7] P. Li and Q. Wang. Local log-Euclidean covariance matrix (L2ECM) for image representation and its applications. In *European Conference on Computer Vision (ECCV)*, pages 469–482, 2012. 2

[8] S. Liao and S. Z. Li. Efficient PSD constrained asymmetric metric learning for person re-identification. In *The IEEE Conference on Computer Vision (ICCV)*, pages 3685–3693, 2015. 2

[9] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *IEEE International Conference on Image Processing (ICIP)*, pages 3567–3571, 2013. 2

[10] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by Fisher vectors for person re-identification. In *European Conference on Computer Vision (ECCV) Workshop*, pages 413–422, 2012. 2

[11] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014. 2

[12] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2666–2672, 2012. 2

[13] S. Paisitkriangkrai, C. Shen, and A. van den Hengel. Learning to rank in person re-identification with metric ensembles. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1846–1855, 2015. 2

[14] S. Pedagadi, J. Orwell, S. A. Velastin, and B. A. Boghossian. Local Fisher discriminant analysis for pedestrian re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3325, 2013. 2

[15] G. Serra, C. Grana, M. Manfredi, and R. Cucchiara. Covariance of covariance features for image classification. In *Proceedings of International Conference on Multimedia Retrieval (ICMR)*, page 411. ACM, 2014. 2

[16] G. Serra, C. Grana, M. Manfredi, and R. Cucchiara. GOLD: Gaussians of local descriptors for image representation. *Computer Vision and Image Understanding*, 134:22–32, 2015. 1

[17] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *European Conference on Computer Vision (ECCV)*, pages 589–600, 2006. 1

[18] A. Vedaldi and B. Fulkerson. Vlfeat – an open and portable library of computer vision algorithms. In *ACM International Conference on Multimedia (ACMMM)*, pages 1469–1472, 2010. 2

[19] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, 2009. 2

[20] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision (ECCV)*, pages 1–16, 2014. 2

[21] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision (ECCV)*, pages 536–551, 2014. 2

[22] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2528–2535, 2013. 2

[23] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3586–3593, 2013. 2

[24] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 144–151, 2014. 2

[25] W. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 649–656, 2011. 2