Person Re-identification by Local Maximal Occurrence Representation and Metric Learning

Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

1 Motivation

Two fundamental problems are critical for person re-identification, feature representation and metric learning.

- Though several descriptors have been proposed for person re-identifi -cation, they are still facing difficulties in invariant representation of a person under illumination variations and viewpoint changes.
- For metric learning, most existing algorithms apply PCA for dimension reduction, and then perform metric learning on the PCA subspace. However, this two-stage processing may not be optimal, because samples of different classes may already be cluttered in the PCA subspace.

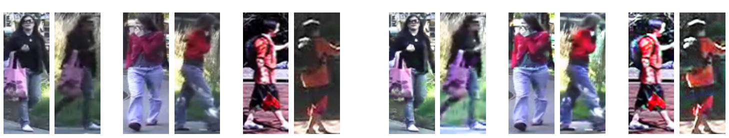
In this paper, we propose

- An efficient feature representation called Local Maximal Occurrence (LOMO), which is robust against viewpoint changes and illumination variations.
- A subspace and metric learning method called Cross-view Quadratic Discriminant Analysis (XQDA), which learns a discriminant low dimensional subspace by cross-view quadratic discriminant analysis, and simultaneously, a QDA metric is learned on the derived subspace.

2 Local Maximal Occurrence Representation

2.1 Role of Retinex

This study shows that applying the Retinex transform helps to reduce the illumination variations between images of the same person. See Fig. 1.



(a) (b)

Figure 1: (a) Example pairs of images from the VIPeR database. (b) Processed images in (a) by Retinex.

The Retinex transform also helps to improve the person re-identification performance, as shown in Fig. 2 (b).

2.2 Dealing with Viewpoint Changes

To address viewpoint changes, we maximize the local occurrence of each pattern among the same horizontal subwindows, as shown in Fig. 2 (a). The effect of this operation can be seen in Fig. 2 (b). Fig. 3 (a) shows that the proposed LOMO feature outperforms other existing features.

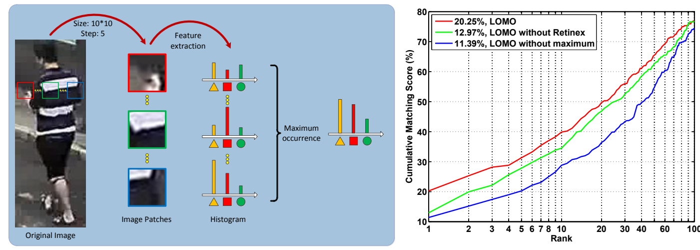
3 XQDA

We learn a subspace W, and a distance function simultaneously:

$$d_W(\mathbf{x},\mathbf{z}) = (\mathbf{x}-\mathbf{z})^T W(\Sigma_I'^{-1} - \Sigma_E'^{-1})W^T(\mathbf{x}-\mathbf{z}), \quad (1)$$

This is further formulated as a Generalized Rayleigh Quotient, and a closed-form solution can be obtained by the generalized eigenvalue decomposition.

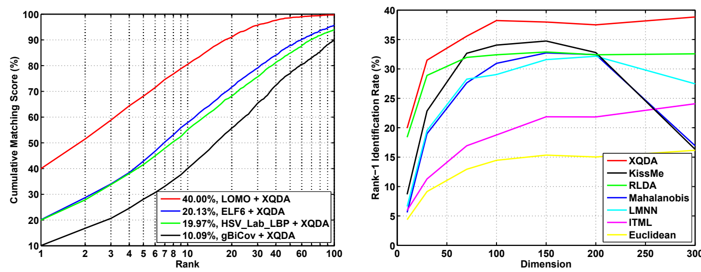
This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) (b)

Figure 2: (a) Illustration of the LOMO feature extraction method. (b) Comparison of the LOMO feature with and without Retinex and the local maximal occurrence operation by the Cosine similarity measure on VIPeR.

We also present a practical computation method for XQDA, as well as its regularization and dimension selection. Fig. 3 (b) shows that XQDA outperforms other existing metric learning methods.

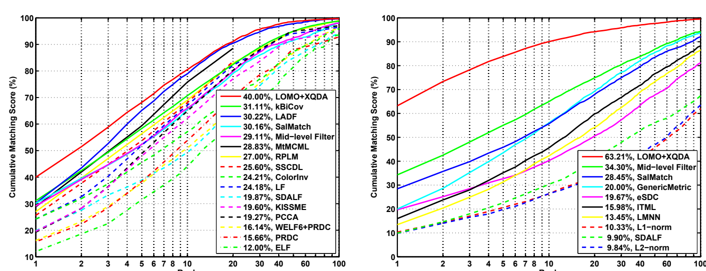


(a) (b)

Figure 3: Performance comparison on the VIPeR database. (a) LOMO and other three features, ELF6, HSV+Lab+LBP, and gBiCov, with the same XQDA. (b) XQDA and other metric learning algorithms with the same LOMO.

4 Experiments

Experiments show that the proposed method improves the state-of-the-art rank-1 identification rates by 2.2%, 4.88%, 28.91%, and 31.55%, respectively, on four challenging person re-identification databases, VIPeR, QMUL GRID, CUHK Campus, and CUHK03. See Fig. 4.



(a) (b)

Figure 4: (a) Comparison of the state of the art with CMC curves and rank-1 identification rates on (a) the VIPeR database (P=316) and (b) the CUHK Campus database (P=486, M=2).

The MATLAB codes of LOMO and XQDA (available at http://www.cbsr.ia.ac.cn/users/scliao/projects/lomo_xqda/) run efficiently: 0.012 seconds per image by LOMO, and 1.86 seconds for training with 316×2 samples by XQDA.