Kernel Partial Least Squares for Person Re-Identification

Raphael Prates, Marina Oliveira and William Robson Schwartz
Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
pratesufop@gmail.com, marinasantoli0308@gmail.com, william@dcc.ufmg.br

Abstract

Person re-identification (Re-ID) keeps the same identity for a person as he moves along an area with non-overlapping surveillance cameras. Re-ID is a challenging task due to appearance changes caused by different camera viewpoints, occlusion and illumination conditions. While robust and discriminative descriptors are obtained combining texture, shape and color features in a high-dimensional representation, the achievement of accuracy and efficiency demands dimensionality reduction methods. At this paper, we propose variations of Kernel Partial Least Squares (KPLS) that simultaneously reduce the dimensionality and increase the discriminative power. The Cross-View KPLS (X-KPLS) and KPLS Mode A capture cross-view discriminative information and are successful for unsupervised and supervised Re-ID. Experimental results demonstrate that X-KPLS presents equal or higher matching results when compared to other methods in literature at PRID450S.

Figure 1. Example of images captured at camera A (first row) and B (last row), in VIPeR dataset. Individuals 4 and 5 are similar at camera A and B. However, it is not common. For instance, individuals 1a, 2a and 3a are similar at camera A, while at camera B they look quite different as a consequence of self-occlusion (1b) and illumination changes (2b).

1. Introduction

Person re-identification (Re-ID) focus on the maintenance of a global identity for an individual as he moves along a network of surveillance cameras without intersection of field-of-views (FOV). Re-ID has an important role in security applications because it allows the observation of subject's actions in a large area.

To use installed infrastructure of surveillance cameras, most of the approaches employ low-resolution images where biometric cues, such as face and iris, are erroneous. Thus, they perform person Re-ID using appearance models learned using clothes information, which consists in a short-time signature. Differently, gait relates to the way a person walks and is a long-time data that can be captured at low-resolution cameras without subject collaboration [16]. However, gait descriptors present a high level of intra-class variation due physiological and external factors [11]. Therefore, designing discriminative and robust features for person Re-ID is still an unsolved problem.

In the common Re-ID scenario, we want to match a probe image captured by a surveillance camera A with a gallery of known individuals, which corresponds to persons previously captured by a surveillance camera B. It is a challenge task because changes caused by pose, illumination, occlusion and camera viewpoints make two different persons look more similar than the same person captured by two distinct cameras (see Fig. 1). As a matter of fact, the direct comparison of probe and gallery images without any previous learning from labelled data (called unsupervised Re-ID [7, 29, 12, 33]) results in a low matching performance (see Table 1). Therefore, some works have used labelled image pairs from both cameras to learn a cross-view matching[1] function, which is known as supervised Re-ID [14, 21, 2, 1, 13, 18, 31].

One possible solution to obtain more robust descriptors corresponds to combine multiple cues at high-dimensional feature vectors. It is a conventional strategy applied when individual's information is available as a single image (single-shot) [17, 5, 7], multiple images (multi-shot) [25, 17, 5, 3, 7], videos (video-based) [15] or complementary

[1] We use the term cross-view matching function to denote methods that are robust to different camera conditions.

sensors, such as thermal [19] and depth [19, 20] cameras.

The combination of multiple cues in high-dimensional feature descriptors increases the Re-ID performance because they can capture subtle characteristics that are simultaneously discriminative and robust to camera transition. However, machine learning methods, such as SVM [6] and the distance metric learning KISSME [13], may not be directly applied at the high-dimensional space due the increased computational cost and reduced performance [27]. In addition, real-world applications require person Re-ID to work in real-time to assist the security personnel. Therefore, we need to device methods that reduce the dimensionality of our data while maintaining its discriminative power to achieve accuracy and efficiency requirements.

Two classical dimensionality reduction methods are PCA and LDA [6]. However, PCA does not consider class labels to compute the low-dimensional representation and LDA estimates between and within classes scatter matrices, which requires sufficient number of samples at each class (at least more than one) and is not suitable for person Re-ID. For instance, LDA will not work in the single-shot scenario, which is the most common at Re-ID literature. Therefore, Re-ID demands a class-aware dimensionality reduction method that does not assume a substantial number of samples at each class.

Re-ID is an ideal scenario for the statistical method Partial Least Squares (PLS) [32]. PLS models the relation between observed variables and responses (e.g. class labels) by computing latent vectors that maximize the correlation while preserving the variance of both sets, making PLS useful for dimensionality reduction, regression and classification problems [4]. As a dimensionality reduction tool, PLS has the advantage of considering the class membership information to find discriminative projections and performs well even in the single-shot scenario.

Some previous works approached Re-ID problem using PLS [27, 21]. Schwartz and Davis [27] computed a discriminative appearance-model for each gallery image using PLS in a one-against-all (PLS-OAA) scheme. To match probe and gallery images, they compute the distance between their latent scores. However, the direct comparison of probe and gallery images disregards the camera transition, one of the main problems at person Re-ID. Prates and Schwartz [21] addressed this problem using labelled image pairs from both cameras (training set) to determine prototypes, which are used to learn a PLS one-against-all model. They claim that if an individual at camera B is similar to the gallery image, its corresponding image at camera A (prototypes) may be similar to the probe. That is a strong assumption because, at real-world scenarios, two similar persons at camera A can be different at camera B (see Fig. 1). Therefore, we believe that PLS has more to contribute with person Re-ID.

Recently, Rosipal and Trejo [24] devised a kernel PLS (KPLS) algorithm to construct a nonlinear regression in the high-dimensional feature space. It is a general framework that involves only linear algebra as the linear PLS, with the exception of kernel matrix computation. KPLS has been successfully applied at computer vision problems such as head pose estimation [10], face recognition [30] and age estimation, gender classification, and ethnicity estimation [9].

The outperforming results of KPLS at related computer vision tasks and the nonlinearity of appearance changes of the same individual at two distinct camera views (see Fig. 1) inspired us to address person Re-ID using a nonlinear Kernel PLS (KPLS) model. To the best of our knowledge, we are the first to address person re-identification using KPLS. Furthermore, we show that KPLS is suitable for both Re-ID scenarios: unsupervised and supervised Re-ID.

In the unsupervised Re-ID, we learn a nonlinear regression model using KPLS that presents improved results when compared to linear PLS-OAA model [27]. For instance, while the linear PLS obtains a rank-1 matching performance of 23.4%, KPLS reaches 40.3% at the PRID450S dataset.

In the supervised Re-ID setting, we propose two KPLS approaches: *Cross-View Kernel PLS (X-KPLS)* and *KPLS Mode A*. *X-KPLS* represents probe and gallery images accordingly with the similarity to training images at its respective camera using a nonlinear regression model. Differently, *KPLS Mode A* finds a common low-dimensional space to maximize similarity between image pairs captured from both cameras. Thus, it is possible to directly compare probe and gallery images at this learned latent space. Experimental results demonstrate that both methods are suitable for person Re-ID and capture cross-view discriminative information. For instance, *X-KPLS* and *KPLS Mode A* reach rank-1 matching performance of 52.8% (33.1%) and 51.5% (31.6%), respectively, which is comparable with state-of-the-art at PRID450S (VIPER) dataset.

2. Kernel Partial Least Squares

We use the following notation in the description. Bold lower-case letters denote vectors and bold upper-case letters denote matrices (e. g., $\mathbf{z}$ and $\mathbf{Z}$, respectively).

Despite being widely used at chemometrics [32], only recently PLS has attracted the attention of computer vision researches [28]. Therefore, we will initially provide an interpretation of PLS to simplify the description of KPLS, which corresponds to its kernel-based variant [24].

PLS models the relation between input variables and responses using score vectors by decomposing the zero-mean matrices of input $\mathbf{X} \in \mathbb{R}^{n \times m}$ and response $\mathbf{Y} \in \mathbb{R}^{n \times q}$ into

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^\top + \mathbf{F} \quad \text{and} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^\top + \mathbf{G}, \end{aligned} \tag{1}$$

where matrices $\mathbf{T}$ and $\mathbf{U} \in \mathbb{R}^{n \times f}$ contain the computed score vectors and the number of factors f is a positive integer. $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{m\times f}$ are the loadings and $\mathbf{F}$ and $\mathbf{G} \in \mathbb{R}^{n\times m}$ are the residuals. In its classical formulation, PLS estimates score vectors $\mathbf{t}$ and $\mathbf{u}$ by means of weight vectors $\mathbf{w}$ and $\mathbf{c}$ such that

$$[cov(\mathbf{t}, \mathbf{u})]^2 = \max_{|\mathbf{w}|=|\mathbf{c}|=1} [cov(\mathbf{Xw}, \mathbf{Yc})]^2, \tag{2}$$

where $cov(\mathbf{t}, \mathbf{u})$ is the sample covariance between $\mathbf{t}$ and $\mathbf{u}$. This problem is solved by Wold [32] using the nonlinear iterative partial least squares (NIPALS) method, a robust procedure to solve singular value decomposition problems.

Rosipal and Trejo [24] proposed a modification on the classical NIPALS that instead of normalizing weight vectors $\mathbf{w}$ and $\mathbf{c}$, normalizes score vectors $\mathbf{u}$ and $\mathbf{t}$, as described in Algorithm 1. After a random initialization of score vector $\mathbf{u}$, it iteratively extracts score vectors $\mathbf{u}$ and $\mathbf{t}$ and deflates matrices $\mathbf{X}$ and $\mathbf{Y}$ by its rank one approximations.

The score vectors $\mathbf{t}$ and $\mathbf{u}$ for each iteration are placed on the columns of matrices $\mathbf{T}$ and $\mathbf{U}$ (Eq. 1) and the final regression of PLS model can be written as

$$\begin{aligned} &\mathbf{Y} = \mathbf{XB} + \mathbf{H}, \quad \text{where} \\ &\mathbf{B} = \mathbf{X}^\top \mathbf{U}(\mathbf{T}^\top \mathbf{X}\mathbf{X}^\top \mathbf{U})^{-1}\mathbf{T}^\top \mathbf{Y}. \end{aligned} \tag{3}$$

Algorithm 1: Partial Least Squares.

input : X, Y matrices and the number of factors (f)

1 randomly initialize $\mathbf{u}$
2 **for** *i=1* **to** f **do**
3   **while** ***u*** *do not converge* **do**
4         $\mathbf{w} = \mathbf{X}^\top \mathbf{u}$
5         $\mathbf{t} = \mathbf{Xw}, \quad \mathbf{t} \leftarrow \frac{\mathbf{t}}{\|\mathbf{t}\|}$
6         $\mathbf{c} = \mathbf{Y}^\top \mathbf{t}$
7         $\mathbf{u} = \mathbf{Yc}, \quad \mathbf{u} \leftarrow \frac{\mathbf{u}}{\|\mathbf{u}\|}$
8   **end**
9     $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{tt}^\top \mathbf{X}, \quad \mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{tt}^\top \mathbf{Y}$
10 **end**

Assuming a nonlinear transformation of input variables to a feature space $\mathcal{F}$, i.e, $\mathbf{\Phi} : x_i \in \mathbb{R}^m \rightarrow \mathbf{\Phi}(x_i) \in \mathcal{F}$, KPLS constructs a linear regression model in $\mathcal{F}$ that corresponds to a nonlinear regression at the original space $\mathbb{R}^m$. Replacing the occurrences of $\mathbf{X}$ by $\mathbf{\Phi}$, which is the mapped matrix into feature space $\mathcal{F}$, we combine lines 4 and 5 of Algorithm 1 to obtain the cross-product $\mathbf{\Phi}\mathbf{\Phi}^\top$. Using the "kernel trick", we avoid the explicitly mapping of data to a high-dimensional space substituting this cross-product by $K_x = \mathbf{\Phi}\mathbf{\Phi}^\top$, where $K_x \in \mathbb{R}^{n\times n}$ is the *kernel Gram matrix*.

In Algorithm 2, we present the Kernel PLS method proposed in [24]. Note that they do not perform a nonlinear map from $\mathbf{Y}$ onto a high-dimensional space, since they want to predict $\mathbf{Y}$ from $\mathbf{K}_x$. For instance, in classification problems, the response matrix $\mathbf{Y}$ assumes the form

Algorithm 2: Kernel Partial Least Squares (KPLS).

input : $\mathbf{K}_x$, Y matrices and the number of factors (f)

randomly initialize $\mathbf{u}$
for *i=1* **to** f **do**
    while ***u*** *do not converge* **do**
        $\mathbf{t} = \mathbf{K}_x \mathbf{w}, \quad \mathbf{t} \leftarrow \frac{\mathbf{t}}{\|\mathbf{t}\|}$
        $\mathbf{c} = \mathbf{Y}^\top \mathbf{t}$
        $\mathbf{u} = \mathbf{Yc}, \quad \mathbf{u} \leftarrow \frac{\mathbf{u}}{\|\mathbf{u}\|}$
    end
    $\mathbf{K}_x \leftarrow (\mathbf{I} - \mathbf{tt}^\top)\mathbf{K}_x(\mathbf{I} - \mathbf{tt}^\top)$
    $\mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{tt}^\top \mathbf{Y}$
end

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_C} & \mathbf{0}_{n_C} & \cdots & \mathbf{1}_{n_C} \end{bmatrix}, \tag{4}$$

where scalar C represents the number of classes and n_i indicates the number of samples in class C_i. $\mathbf{1}_{n_i}$ and $\mathbf{0}_{n_i}$ denote a $n_i \times 1$ vector of ones or zeros, respectively.

Similarly to PLS, in KPLS we can predict the responses as $\mathbf{Y} \approx \Phi\mathbf{B}$. Thus, for the training and testing data, we obtain the responses $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_t$, respectively, as

$$\begin{aligned} &\hat{\mathbf{Y}} = \mathbf{TT}^\top \mathbf{Y}, \text{ and} \\ &\hat{\mathbf{Y}}_t = \mathbf{K}_t \mathbf{U}(\mathbf{T}^\top \mathbf{KU})^{-1}\mathbf{T}^\top \mathbf{Y}. \end{aligned} \tag{5}$$

PLS-based methods are useful to learn a common subspace where data from different blocks of variables (e.g, camera views) are successfully matched. For instance, PLS Mode A is a symmetric variant of PLS suitable to learn latent linear relations between two block of variables [23]. However, in many real-world problems (e.g., person Re-ID), the feature transformation between blocks of data demands nonlinear approaches. Therefore, we propose a kernel extension of PLS Mode A that nonlinearly maps both $\mathbf{X}$ and $\mathbf{Y}$ to high-dimensional space.

In Section 3, we use KPLS in unsupervised Re-ID problem. Furthermore, we present two extensions of KPLS to handle cross-view matching: *Cross-View Kernel PLS (X-KPLS)* and *KPLS Mode A*.

3. Proposed Approaches

In this work, we consider the single-shot scenario (i.e., there is a single image taken from camera A and one taken from camera B). We use matrices $\mathbf{T}^a \in \mathbb{R}^{n\times m}$ and $\mathbf{T}^b \in \mathbb{R}^{n\times m}$ to store the n training images from camera A and B, respectively. We represent the i*th* training image from camera A and B as row vectors $\mathbf{t}_i^a$ and $\mathbf{t}_i^b \in \mathbb{R}^m$, respectively, where m denote the dimension of the feature space. Without loss of generality, we assume that l testing images from camera A constitute the probe set $\mathbf{P} \in \mathbb{R}^{l \times m}$ and l testing images from camera B represent the gallery set $\mathbf{G} \in \mathbb{R}^{l \times m}$.

In the following sections, we address person Re-ID using variations of Kernel PLS. Section 3.1 discusses the direct application of KPLS at the person Re-ID problem. Despite of interesting results, this approach cannot consider the camera transition problem. Therefore, Section 3.2 presents a formulation of KPLS that incorporates cross-view discriminative information by using labelled image pairs from the training set, which we named *Cross-View Kernel PLS (X-KPLS)*. Furthermore, Section 3.3 presents the *KPLS Mode A*, a different formulation that, instead of performing regression, learns a common subspace where the direct comparison of probe and gallery images results in high matching performance.

3.1. KPLS-based unsupervised Re-ID

Unsupervised Re-ID assumes that only probe and gallery images are available. Therefore, we need to directly match probe and gallery images disregarding the camera transition problem. However, instead of comparing probe and gallery images directly, we compute a signature using the KPLS regression model and consider the entire gallery set.

To represent the probe and gallery images using the entire gallery set, we define as the element i, j of *Kernel Gram matrix*, represented by $\mathbf{K}_x(i, j)$, the computation of a kernel function at gallery images $\mathbf{g}_i, \mathbf{g}_j \in \mathbf{G}$. In addition, we define the regression responses using Equation 4, where C and n_C are equal to l and 1, respectively, since we have l gallery images and consider the single-shot scenario. Thus, we compute the score vectors $\mathbf{T}$ and $\mathbf{U}$ using Algorithm 2 and the gallery signatures are regression responses $\hat{\mathbf{Y}}$ (Eq. 5).

For the probe image $\mathbf{p}_i \in \mathbf{P}$, we define the row-vector $\mathbf{k}_t^i \in \mathbb{R}^{1 \times n}$, where an element $\mathbf{k}_t^i(j)$ corresponds to the similarity between probe image i and gallery image j, and compute the response $\hat{\mathbf{Y}}_t$ according to Equation 5. It is important to highlight that in Equation 5, we represent a more general case using matrix $\mathbf{K}_t$ but, here, we replace it by a row-vector $\mathbf{k}_t^i$ because at the common person Re-ID scenario, we only have one probe image at each query.

The regression responses $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_t$ (signatures) correspond to the similarity of gallery and probe images, respectively, with respect to the entire gallery set. We assume that it provides useful contextual information that we cannot have when matching them directly. Therefore, we compute the similarity between probe and gallery images using the cosine similarity of its respective signatures ($\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_t$).

3.2. Cross-view Kernel PLS (X-KPLS)

In *X-KPLS*, we use the training images available in cameras A and B to incorporate cross-view discriminative information. Similarly to Prates and Schwartz [21], we compare probe and gallery images indirectly using training images at its respective cameras. However, instead of finding prototypes, we use a nonlinear regression model (KPLS).

Our goal is to represent a probe image $\mathbf{p}_i \in \mathbf{P}$ and a gallery image $\mathbf{g}_j \in \mathbf{G}$ using training images at their respective cameras, $\mathbf{T}^a$ and $\mathbf{T}^b$. Thus, we first define the matrix $\mathbf{K}$, whose element i, j is the result of a kernel function using as input the feature descriptors $\mathbf{t}_i^a$ and $\mathbf{t}_j^a$. Then, we compute the probe's signature $\hat{\mathbf{Y}}_t^{\mathbf{p}_i}$ (Eq. 5) using matrices $\mathbf{T}$ and $\mathbf{U}$ obtained using Algorithm 2 and the row-vector $\mathbf{k}_t^i \in \mathbb{R}^{1 \times n}$, whose element i, j corresponds to the kernel function applied using probe image $\mathbf{p}_i$ and training image $\mathbf{t}_j^a$. Similarly, we compute the signature for jth gallery image $\hat{\mathbf{Y}}_t^{\mathbf{g}_j}$ using the training images at camera B ($\mathbf{T}^b$).

We assume that signatures $\hat{\mathbf{Y}}_t^{\mathbf{p}_i}$ and $\hat{\mathbf{Y}}_t^{\mathbf{g}_j}$ correspond to an indirect form of computing the similarity between a probe image $\mathbf{p}_i$ and gallery image $\mathbf{g}_j$. It means that similarities between gallery and training images at camera B will also occur at probe and training images at camera A. Therefore, we define the similarity between the probe image $\mathbf{p}_i$ and the gallery image $\mathbf{g}_j$ as the cosine between $\hat{\mathbf{Y}}_t^{\mathbf{p}_i}$ and $\hat{\mathbf{Y}}_t^{\mathbf{g}_j}$.

3.3. Kernel PLS Mode A

Algorithm 3: Kernel Partial Least Squares Mode A.

input : $\mathbf{K}_x$, $\mathbf{K}_y$ matrices and the number of factors

randomly initialize $\mathbf{u}$
for *i=1* **to** *f* **do**
  while ***u*** *do not converge* **do**
    $\mathbf{t} = \mathbf{K}_x \mathbf{u}, \quad \mathbf{t} \leftarrow \frac{\mathbf{t}}{||\mathbf{t}||}$
    $\mathbf{u} = \mathbf{K}_y \mathbf{t}, \quad \mathbf{u} \leftarrow \frac{\mathbf{u}}{||\mathbf{u}||}$
  end
  $\mathbf{K}_x \leftarrow (\mathbf{I} - \mathbf{t}\mathbf{t}^\top)\mathbf{K}_x(\mathbf{I} - \mathbf{t}\mathbf{t}^\top)$
  $\mathbf{K}_y \leftarrow (\mathbf{I} - \mathbf{u}\mathbf{u}^\top)\mathbf{K}_y(\mathbf{I} - \mathbf{u}\mathbf{u}^\top)$
end

In *Kernel PLS Mode A*, instead of performing a regression, our goal is to nonlinearly map images from camera A and B to a low-dimensional common subspace where images of the same person have highly correlated features.

To nonlinearly correlate feature descriptors of the same person from camera A and B, we define the matrices $\mathbf{K}_x$ and $\mathbf{K}_y$ using labelled image pairs from camera A and B. Thus, each element *i,j* from matrix $\mathbf{K}_x$ corresponds to a kernel function applied at training images $\mathbf{t}_i^a$, $\mathbf{t}_j^a \in \mathbf{T}^a$. Similarly, we define matrix $\mathbf{K}_y$ at element i,j as a kernel function applied at $\mathbf{t}_i^b$, $\mathbf{t}_j^b \in \mathbf{T}^b$. Finally, we compute the score vectors **T** and **U** accordingly to Algorithm 3.

To project new samples onto the learned low-dimensional space, we define projection matrices $\mathbf{W}_x$ and $\mathbf{W}_y \in \mathbb{R}^{n\times d}$ that relate score vectors **T** and **U** with matrices $\mathbf{K}_x$ and $\mathbf{K}_y$, respectively, such that

$$\begin{aligned}\mathbf{T} &= \mathbf{K}_x\mathbf{W}_x \text{ and}\\ \mathbf{U} &= \mathbf{K}_y\mathbf{W}_y,\end{aligned} \tag{6}$$

where projection matrices $\mathbf{W}_x$ and $\mathbf{W}_y$ are computed as

$$\begin{aligned}\mathbf{W}_x &= \mathbf{T}(\mathbf{T}^\top\mathbf{K}_x\mathbf{T})^{-1} \text{ and}\\ \mathbf{W}_y &= \mathbf{U}(\mathbf{U}^\top\mathbf{K}_y\mathbf{U})^{-1}.\end{aligned} \tag{7}$$

For a gallery image $\mathbf{g}_i \in \mathbf{G}$, we first compute its row-vector representation $\mathbf{k}_g^i \in \mathbb{R}^{1\times n}$, whose element $\mathbf{k}_g^i(j)$ corresponds to the kernel function applied using feature descriptors $\mathbf{g}_i$ and $\mathbf{t}_j^b \in \mathbf{T}^b$. Then, we obtain its low-dimensional representation $\hat{\mathbf{g}}_i$ such that

$$\hat{\mathbf{g}}_i = \mathbf{k}_g^i\mathbf{W}_y. \tag{8}$$

Similarly, for a given probe image $\mathbf{p}_i \in \mathbf{P}$, we compute its vector $\mathbf{k}_p^i \in \mathbb{R}^{1\times n}$ using training images at camera A ($\mathbf{T}^a$) and obtain its representation at a low-dimensional space as

$$\hat{\mathbf{p}}_i = \mathbf{k}_p^i\mathbf{W}_x. \tag{9}$$

Finally, we use the cosine between $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{g}}_i$ to compute the similarity between $\mathbf{p}_i$ and $\mathbf{g}_i$.

4. Experimental Results

In this section, considering supervised and unsupervised Re-ID scenarios, we evaluate the proposed variations of KPLS on two widely used datasets for the Re-ID problem (VIPeR and PRID450S datasets). First, we present these datasets and the evaluation protocol (Section 4.1). Then, we present a comparison between the proposed methods with state-of-the-art approaches in unsupervised (Section 4.2) and supervised (Section 4.3) person Re-ID problems.

4.1. Datasets and Evaluation Protocol

VIPeR Dataset[2] [8]. VIPeR dataset contains 632 labelled image pairs captured by two different outdoor cameras. Each camera captures a single image of each subject (single-shot) and the images are normalized to 128×48 pixels. VIPeR is the most used dataset for supervised and unsupervised Re-ID. The main challenges are related to viewpoints, illumination and low-resolution images (see Fig. 1).

[2] Available at: https://vision.soe.ucsc.edu/projects

PRID 450S Dataset[3] [22]. PRID 450S is a recently released dataset with more realistic images. It consists in 450 images pairs of pedestrians captured by two non-overlapping cameras. Each subject appears in single image at each camera (single-shot). The main challenges are related to changes in viewpoint, pose as well as significant differences in background and illumination.

Experimental Setup. Similarly to other works in the literature, we achieve more stable results using ten random partitions of the dataset in training and testing subsets of equal sizes and reporting the mean values. In the testing subset, images from one camera are considered as gallery and images from the other as probe. The results are reported using the matching performance at the *top-r* positions, which corresponds to the number of probe individuals correctly identified at the first r returned gallery individuals.

Feature Extraction. To represent images on the VIPeR dataset, we use the 5138-dimensional feature descriptor proposed in [14], which is a histogram-based combination of color, edges and texture descriptors. Differently, we obtained improved results at PRID450S using the 35024-dimensional combination of hand-crafted and Convolutional Neural Network (CNN) features described in [26].

Parameter Settings. We adjust the number of factors (f) of Kernel PLS methods using random splits between training and testing sets.As a result, we set f to 300, 100 and 125 for KPLS, *X-KPLS* and *KPLS Mode A*, respectively. Furthermore, we compute the *kernel Gram matrices* using exponential χ^2 kernel function, as described in [14].

4.2. Unsupervised Re-ID

Compared Methods. We compare our approach with a baseline method (PLS-OAA) [27], the SDALF [7] method, salience learning method SDC [33], the semantic attribute learning proposed by Shi et *al.* [29] and the unsupervised dictionary learning method introduced by Kodirov et *al.* [12] (for a fair comparison, we use the code provided by the Kodirov et *al.* [12] and execute with the features described in Section 4.1 for PRID450S).

Discussion. The obtained results at VIPeR and PRID450S datasets are presented in Tables 1 and 2. Experimental results considering both datasets demonstrate that the Kernel PLS presents a significant improvement when compared with the baseline PLS-OAA method. Furthermore, we obtain state-of-the-art results at PRID450S, particularly at the initial ranking positions. However, these improvements are not consistent when evaluated at VIPeR dataset. We attribute these deviations to the fact that our method assumes consistent camera transitions, which is not true for the VIPeR dataset.

[3] Available at: https://lrs.icg.tugraz.at/download.php

Table 1. Unsupervised Re-ID Viper: top ranked CMC results.

Method	Viper (p=316)				
	r = 1	r = 5	r = 10	r = 20	r = 30
PLS-OAA [27]	10.4	25.9	35.3	46.0	53.4
SDALF [7]	19.9	40.0	49.4	65.7	75.6
SDC [33]	25.1	44.9	56.3	70.9	-
Shi et *al.* [29]	27.7	**55.3**	**68.3**	**79.7**	-
Kodirov et *al.* [12]	**29.6**	54.8	64.8	77.3	-
KPLS	22.9	45.0	56.9	67.5	73.3

Table 2. Unsupervised Re-ID PRID450S: top ranked CMC results.

Method	PRID450S (p=225)				
	r = 1	r = 5	r = 10	r = 20	r = 30
SDALF [7]	17.4	30.9	40.8	55.2	-
PLS-OAA [27]	23.4	48.3	61.3	74.2	81.1
SDC [33]	23.7	38.4	46.1	58.5	-
Shi et *al.* [29]	28.5	48.9	59.6	71.3	-
Kodirov et *al.* [12]	39.4	61.6	70.6	78.5	83.8
KPLS	**40.3**	**63.1**	**71.6**	**81.3**	**87.2**

4.3. Supervised Re-ID

Compared Methods. We compare the proposed methods (*X-KPLS* and *KPLS Mode A*) with a supervised PLS approach proposed by Prates and Schwartz [21], the well-known distance metric learning methods PRDC [31], KISSME [13] and EIML [18], and subspace learning approaches RCCA [1], ROCCA [2] and KCCA [14]. For KCCA [14] in PRID450S, we use the code provided by the authors and the features described in Section 4.1, which we denote as KCCA + CNN Features. Some approaches are missing in Table 2 because they neither provide their code nor results for the dataset PRID450S.

Discussion. In Tables 3 and 4, we present the obtained results in VIPeR and PRID450S datasets, respectively. These results show that *X-KPLS* presents improved performance when compared to the *KPLS Mode A* and the baseline PLS method [21]. We attribute this improvement to the representation of probe and gallery images using the entire training set at their respective cameras. Differently, *KPLS Mode A* assumes that exists a common subspace where the comparison between different cameras is successful while Prates and Schwartz [21] restrict the analyze to small number of training samples (prototypes). That are strong assumptions that do not hold for complex camera transitions.

Considering the VIPeR dataset, the comparison between *X-KPLS* and *KPLS Mode A* show improved results when compared with metric learning methods and comparable results with other subspace learning methods. KCCA [14] presents better matching performance when compared with our both approaches. However while KCCA uses more than 300 dimensions for the subspace, our method requires 100 and 125 for *X-KPLS* and *KPLS Mode A*, respectively, making our method more efficient. In PRID450S, *X-KPLS* reaches equal or higher matching performance when compared with KCCA method and requires significantly lower subspace dimensions (or factors).

Table 3. Supervised Re-ID Viper: top ranked CMC results.

Method	Viper (p=316)				
	r = 1	r = 5	r = 10	r = 20	r = 30
PRDC [31]	15.7	38.4	53.9	70.0	-
EIML [18]	22.0	47.5	63.0	78.0	87.0
KISSME [13]	27.0	55.0	70.0	83.0	89.5
RCCA+RD [1]	30.2	60.0	74.7	86.8	-
ROCCA [2]	30.4	-	75.6	86.6	-
Prates and Schwartz [21]	32.9	62.3	78.7	87.8	91.6
KCCA [14]	**37.0**	-	**85.0**	**93.0**	-
KPLS Mode A	31.6	64.0	77.8	88.1	91.8
X-KPLS	33.1	**67.7**	81.0	90.8	**94.2**

Table 4. Supervised Re-ID PRID450S: top ranked CMC results.

Method	PRID450S (p=225)				
	r = 1	r = 5	r = 10	r = 20	r = 30
Prates and Schwartz [21]	29.3	52.5	63.1	75.0	82.1
KISSME [13]	33.0	59.8	71.0	79.0	84.5
EIML [18]	35.0	58.5	68.0	77.0	83.0
KCCA + CNN Features	**52.8**	80.9	89.0	95.1	97.2
KPLS Mode A	51.5	78.6	87.0	93.7	96.0
X-KPLS	**52.8**	**82.1**	**90.0**	**95.4**	**97.3**

5. Conclusions

This paper addressed the person Re-ID problem using the powerful statistical method called Kernel Partial Least Squares (KPLS). To the best of our knowledge, this is the first time that KPLS was employed to solve the Re-ID problem. Furthermore, we presented two extensions of Kernel PLS to incorporate cross-view discriminative information available at labelled image pairs: *Cross-View Kernel PLS (X-KPLS)* and *KPLS Mode A*. The experimental results demonstrated that KPLS is successful for unsupervised and supervised Re-ID problem. For instance, we obtain state-of-the-art results at PRID450S dataset in both scenarios.

Acknowledgments

The authors would like to thank the Minas Gerais Research Foundation – FAPEMIG (Grant APQ-00567-14) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project). Part of the results presented in this paper were obtained through research on a project titled ”HAR-HEALTH: Reconhecimento de Atividades Humanas associadas a Doenças Crônicas”, sponsored by Samsung Eletrônica da Amazônia Ltda. under the terms of Brazilian federal law No. 8.248/91.

References

[1] L. An, M. Kafai, S. Yang, and B. Bhanu. Reference-based person re-identification. In *IEEE AVSS*, pages 244–249, Aug 2013.

[2] L. An, S. Yang, and B. Bhanu. Person re-identification by robust canonical correlation analysis. *Signal Processing Letters, IEEE*, 22(8):1103–1107, Aug 2015.

[3] L. Bazzani, M. Cristani, A. Perina, and V. Murino. Multiple-shot person re-identification by chromatic and epitomic analyses. *Pattern Recognition Letters*, 33(7):898–903, 2012.

[4] K. P. Bennett and M. J. Embrechts. An optimization perspective on kernel partial least squares regression. *Nato Science Series*, 3(190):227–250, 2003.

[5] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011.

[6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE CVPR*, pages 2360–2367. IEEE, 2010.

[8] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE PETS*, 2007.

[9] G. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *IEEE CVPR*, pages 657–664. IEEE, 2011.

[10] M. Haj, J. Gonzalez, and L. S. Davis. On partial least squares in head pose estimation: How to simultaneously deal with misalignment. In *IEEE CVPR*, pages 2602–2609. IEEE, 2012.

[11] Y. Iwashita, H. Sakano, and R. Kurazume. Gait recognition robust to speed transition using mutual subspace method. In *Image Analysis and ProcessingICIAP 2015*, pages 141–149. Springer, 2015.

[12] E. Kodirov, T. Xiang, and S. Gong. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification. In *BMVC*, pages 44.1–44.12. BMVA Press, September 2015.

[13] M. Koestinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE CVPR*, 2012.

[14] G. Lisanti, I. M., and A. Del Bimbo. Matching people across camera views using kernel canonical correlation analysis. pages 10:1–10:6. ACM, 2014.

[15] K. Liu, B. Ma, W. Zhang, and R. Huang. A spatio-temporal appearance representation for video-based pedestrian re-identification. In *IEEE ICCV*, pages 3810–3818, Dec 2015.

[16] Z. Liu, Z. Zhang, Q. Wu, and Y. Wang. Enhancing person re-identification by integrating gait biometric. *Neurocomputing*, 168:1144–1156, 2015.

[17] H. Martin, B. Csaba, R. P. M., and B. Horst. Person re-identification by descriptive and discriminative classification. In *Image Analysis*, pages 91–102. Springer, 2011.

[18] H. Martin, R. Peter, and B. Horst. Person re-identification by efficient impostor-based metric learning. In *IEEE AVSS*, pages 203–208. IEEE, 2012.

[19] A. Mogelmose, C. Bahnsen, T. Moeslund, A. Clapes, and S. Escalera. Tri-modal person re-identification with rgb, depth and thermal features. In *IEEE CVPRW*, pages 301 – 307, June 2013.

[20] F. Pala, R. Satta, G. Fumera, and F. Roli. Multi-modal person re-identification using rgb-d cameras. *IEEE TCSVT*, PP(99):1–1, 2015.

[21] R. Prates and W. Schwartz. Appearance-based person re-identification by intra-camera discriminative models and rank aggregation. In *IAPR ICB*, pages 65–72, 2015.

[22] P. R., M. Hirzer, M. Kstinger, C. B., and H. B. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, Springer ACVPR, pages 247–267. 2014.

[23] R. Rosipal and N. Krämer. Overview and recent advances in partial least squares. In *Subspace, latent structure and feature selection*, pages 34–51. Springer, 2006.

[24] R. Rosipal and L. Trejo. Kernel partial least squares regression in reproducing kernel hilbert space. *The Journal of Machine Learning Research*, 2:97–123, 2002.

[25] K. Ryosuke, M. Yasushi, H. Chunsheng, I. Haruyuki, and Y. Yasushi. Person re-identification using view-dependent score-level fusion of gait and color features. In *IAPR ICPR*, pages 2694–2697. IEEE, 2012.

[26] X. L. J. Y. W. Z. S. Wu, Y. Chen. An enhanced deep feature representation for person re-identification. In *IEEE WACV*, March 2016.

[27] W. Schwartz and L. Davis. Learning discriminative appearance-based models using partial least squares. In *SIBGRAPI*, pages 322–329, Oct 2009.

[28] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human detection using partial least squares analysis. In *Computer vision, 2009 IEEE 12th international conference on*, pages 24–31. IEEE, 2009.

[29] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *IEEE CVPR*, pages 4184–4193, 2015.

[30] V. Štruc and N. Pavešić. Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20(1):115–138, 2009.

[31] Z. Wei-Shi, G. Shaogang, and X. Tao. Person re-identification by probabilistic relative distance comparison. In *IEEE CVPR*, pages 649–656. IEEE, 2011.

[32] H. Wold. *Encyclopedia of Statistical Sciences*, volume 6. John Wiley & Sons, 1985.

[33] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *IEEE CVPR*, pages 3586–3593, June 2013.