Single-Shot Person Re-Identification Combining Similarity Metrics and Support Vectors

Anderson Luís Cavalcanti Sales[1], Rafael Henrique Vareto[2], William Robson Schwartz[2], Guillermo Camara Chavez[1]
[1]Universidade Federal de Ouro Preto, Ouro Preto, Minas Gerais, Brazil
[2]Smart Sense Laboratory, Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
{dersonluis, gcamarac}@gmail.com, {rafaelvareto, william}@dcc.ufmg.br

***Abstract*—Person Re-Identification is all about determining a person's entire course as s/he walks around camera-equipped zones. More precisely, person Re-ID is the problem of matching human identities captured from non-overlapping surveillance cameras. In this work, we propose an approach that learns a new low-dimensional metric space in an attempt to cut down multi-camera matching errors. We represent the training and test samples by concatenating handcrafted features. Then, the method performs a two-step ranking using elementary distance metrics and followed by an ensemble of weighted binary classifiers. We validate our approach on CUHK01 and PRID450s datasets, providing only a sample per class for probe and only a sample for gallery (single-shot). According to the experiments, our method achieves CMC Rank-1 results up to 61.1 and 75.4, following leading literature protocols, for CUHK01 and PRID450s, respectively.**

I. Introduction

Person Re-Identification (Re-ID) is the inter-camera human association that tracks individuals through distinct field-of-view (FoV) cameras. More precisely, person Re-ID is the problem of matching human identities captured from non-overlapping surveillance cameras. For a clearer understanding, think of a set of cameras installed in an office building labeled from *A* to *Z*. When a subject walks from camera's *A* FoV to camera's *B* FoV, human re-identification attempts to identify a unified path from discontinuous tracks, tracking the individual's full course on multiple cameras. In traditional person Re-ID methods, given a query image or a collection of images from an unknown subject, also known as probe image, and a gallery set comprised of numerous pictures of known individuals, the idea lies on building a ranked list containing all persons enrolled in the gallery set taking into account their similarity to the unknown probe image.

Because of the increasingly number of camera networks set out in areas such as office buildings, shopping centers, airports, railway stations, person re-identification tasks have earned a significant consideration from the research community. Such surveillance systems generate many hours of videos, making the manual processing impractical. As a consequence, the human supervision of multi-camera videos tend to be inaccurate, time-consuming and exhausting, which critically reduces the monitoring efficiency. However, situation awareness is a fundamental step for qualified security. Hampapur et al. [1] state that intelligent surveillance systems composed by information from diverse strands are key to exercising conscious security in various scenarios these days. They also claim that automatic video analysis is capable of non-intrusively detecting activities and predicting undesirable activities, making the security staff more pro-active [2].

The predominant challenge for person Re-ID is the person's appearance discrepancy over different cameras [3]. For an accurate Re-ID system, satisfactory feature descriptors have to be obtained from visual data under unrestrained circumstances, which people do not cooperate with data acquisition. Many surveillance cameras are not installed properly or do not capture good-quality videos, outputting low frame rate and resolution. In addition, there is a high chance of subjects being fully or partially occluded by objects or by other people. Depending on the requirements, a Re-ID surveillance system must deal with single images per person (*a.k.a.* single-shot task) or take videos as input (*a.k.a.* multi-shot task), indicating that several person-containing frames are available to compound probe and gallery sets. More frames per subject are desirable since more instances of a person are explored to generate better discriminative systems. All these limitations turn person re-identification into a complex problem, especially when there are numerous individuals enrolled in the gallery set as the large number of subjects may lead to specificity loss, escalating the probability of identification inaccuracy.

In defiance of all existing challenges, this work consists of single-shot person Re-ID. We extract two feature representations widely used in the literature to learn a new low-dimensional metric subspace in pursuance of reduced multi-camera matching errors and robust similarity functions. These errors occur frequently in the human re-identification domain mainly due to lighting, pose and viewpoint changes. On the new subspace, we employ a two-stage ranking based on either cosine or Mahalanobis distance [4], [5] and binary classifiers. Initially, the proposed method generates a list of candidates based on one of the previously mentioned distance metrics, where gallery samples that closest match the probe image lead the list. Only individuals on the top of the ranking proceed to the second stage. Influenced by the work of Vareto et al. [6], we learn a small ensemble of binary classifiers, for top candidates only every time there is a probe query in favor of augmenting the method's accuracy. The approach subdivides probe and gallery set samples in overlapping sections so that

each ensemble classifier is dedicated to a segment of the image. The final candidate list is generated using multiple Support Vector Machine (SVM) classification models.

The main hypothesis behind our work lies on two fundamental steps: (i) a constructive algorithm established on distance metrics resulting in a reduced candidate list, and (ii) classification models that are capable of refining the overall performance of the approach. In other words, experiments show that running a collection of classifiers following a straightforward similarity criterion significantly improves the precision of the method.

The most relevant contributions are: (i) the employment of simple distance metrics followed by an ensemble of weighted binary classifiers; (ii) an easy-to-implement algorithm with just a few parameters to be set; and (iii) a low-computational cost method, suited to being deployed on embedded systems, in comparison with deep neural network methods.

The remaining of this paper is organized as follows. Section II contains literature works that are relevant to our approach, Section III further describes our method, Section IV outlines protocols and experiments, and Section V concludes with final remarks.

II. Related Works

Person re-identification (Re-ID) may be considered a challenging problem, even for the human eye. In general, a person's appearance usually varies across all cameras since there are completely distinct illumination and camera viewpoints. For that reason, Re-ID has become an intense researched topic in the last decade [7]–[14].

One of the most important elements for person re-identification is the way color and textural information are characterized. When it comes to surveillance, cameras cover wide areas. As a result, subjects are commonly framed in low resolution and in a mixture of pose variations. This constraining condition is sufficient to make researchers explore alternate methods, which contemplate both color name descriptors and color histograms [15], [16]. There is a large number of researchers working on appearance-based descriptors seeing that these descriptors are responsible for holding appearance information of people's clothes [15]–[18].

Nappi et al. [12] propose a generic semi-supervised approach for face and object re-identification through a modular architecture oriented to online data. Zhao et al. [15] design a method to learn human salience in an unsupervised manner, then they apply patch matching to build a correspondence between image pairs and reduce multi-camera misalignment. Yang et al. [16] propose a new color descriptor combined with a simple metric learning method that enriches the feature representation. Ma et al. [17] come up with a new color representation for person Re-ID, a feature descriptor that not only models color characteristics, but also represents texture and spatial structures. Some researchers prefer to combine different color feature descriptors with some texture descriptors like local binary pattern [19], or even different methods to represent images and signals, for instance, wavelet transforms and filter banks. Matsukawa et al. [20] propose a different feature descriptor based on the pixels' hierarchical covariance. The method describes local regions in an image with a hierarchical Gaussian distribution considering a set of multiple Gaussian distributions where every single Gaussian portrays the form of a local patch. To that end, the descriptor ends up modeling both mean and covariance as the authors have concluded that the mean color in local parts have a major discriminating contribution for matching subjects across cameras. Different from previously mentioned methods, Xiao et al. [21] present a method of "muting" neurons on a domain based dropout algorithm to learn deep feature representations. Furthermore, Varior et al. [22] propose a Siamese CNN for person Re-ID that adaptively boosts local features for enhancing the discriminative capability of the network.

Many approaches encompass appearance-based similarity in order to identify their resemblance and establish a prospect equivalence. Most works from the literature employ low-level feature descriptors containing color and texture information, generally extracted from clothing. Therefore, features obtained from garments tend to be reliable over short-time periods only as individuals usually dress different clothes on adjacent days. As a result, most state-of-the-art works in the literature seek solutions for the short-time period scenario [5], [11], [13], [14], [21]–[24].

Several authors have dedicated their efforts on learning latent metric spaces as they realized it may attain relevant results when combined with distance-based techniques such as nearest neighbors. They learn new metrics by minimizing the intra-class distances as well as maximizing the inter-class distances. In general, metric-learning based methods are limited by its inner information loss and error, caused by the subtraction of misaligned feature vectors.

Li et al. [11] tackle the Re-ID problem through a transfer learning framework where training samples are chosen and re-weighted in accordance with their distance approximation to the query image and a list of candidates. The metric-learning framework learns specific metrics for individual probe-candidate settings. Hirzer et al. [25] developed a method that learns the transition from one camera to another by means of the Mahalanobis metric using pairs of tagged samples from distinct cameras. In addition, Roth et al. [5] evaluate how metric learning the techniques derived from Mahalanobis distance can be applied to single-shot person re-identification, that is, considering individuals with a single image sample. Liao et al. [23] design new techniques for feature representation and metric learning, the most relevant stages in person Re-ID. The descriptor takes into account horizontal occurrences of local features to make a robust description in an attempt to reduce the variability resulted from different viewpoints. The new low-dimensional subspace is built from a cross-view quadratic discriminant analysis, and simultaneously, a quadratic discriminant analysis metric is learned on the derived subspace. In other words, the framework concurrently learns a discriminant subspace and a distance metric so that it is able to pick the optimal dimensionality.

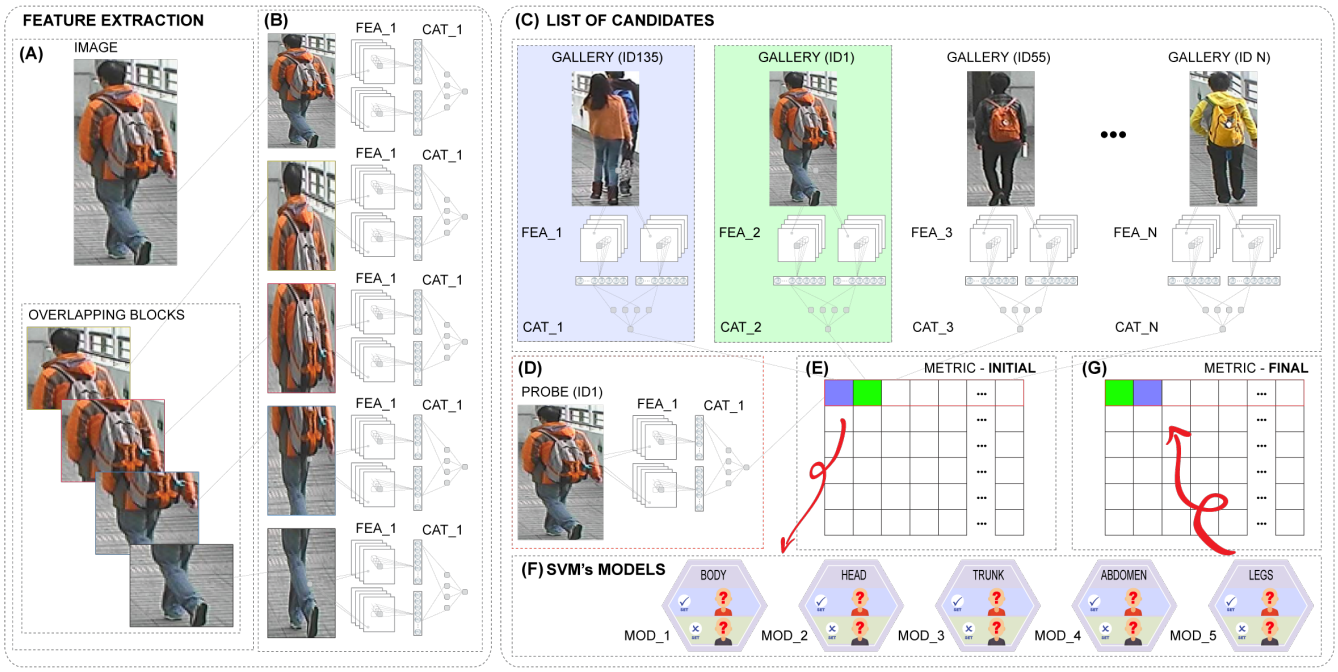

Fig. 1. Illustration of the proposed pipeline: **Sample partition:** (A-B) The original image is divided among four overlapping chunks. LOMO and GOG feature representations are extracted from the initial whole-body image and all four chunks separately. After that, LOMO and GOG features are concatenated into a new elongated feature descriptor, which is used to learn a low-dimensional XQDA representation. **Test procedure:** (C-E) Given a probe p and a full-body gallery sample $g_i \in G$, it computes the similarity distance between p and g_i, resulting in an initial list of candidates. (F-G) For leading candidates, original full-body and derived samples are used to train five SVM models for the sake of improving the method's overall performance.

Prates et al. [26] reach a considerable high matching performance employing kernel PCA, a statistical method that learns a shared subspace and captures most of the variability with just few vector basis. The work engages in a framework that addresses both camera transitions and dimensionality reduction with a reduced number of dimensions. In another work, Prates et al. [27] propose some adaptations for the kernel partial least squares. Partial least squares (PLS) is a statistical technique capable of simultaneously diminishing dimensionality and increasing discerning power. According to the authors, kernel-based PLS captures distinctive information, vital for multi-class problems like person Re-ID. Kernel PCA and PLS are widely used in other areas, such as medicine, for genomic prediction and disease detection [28]–[30]. Even though kernel-derived methods came up with great contribution to re-identification, kernel PLS neglects the fact that cameras are captured on different field of views, which is critical for the Re-ID problem.

As detailed in this section, the vast majority of human re-identification methods initially extract features from a set of cameras. Then, they build a new metric space that feeds machine learning algorithms. Unlike the greater part of works in the literature, we first compute simple cosine distances between a probe and all samples from the gallery set. This comparison produces a list of candidates where most similar enrolled individuals are chosen to estimate Support Vector Machine (SVM) models in pursuance of an enhanced recognition. We detail the proposed approach in the next section.

III. Proposed Approach

As mentioned in Section I, image variations such as lighting and pose changes, among other factors, are responsible for the main difficulties encountered in person re-identification (Re-ID). To mitigate damaging effects resulting from camera capture in non-cooperative surveillance settings, the proposed method combines different feature representations, projecting them onto a new common subspace, later adjusting potential candidate samples to better ranking positions using binary classifiers. In addition, the proposed Re-ID method partitions the original[1] whole-body image into intersecting blocks/chunks. This is an attempt to increase the gallery set size and perform data augmentation since the approach focuses on single-shot tasks. From now on, we refer to both full-body image and its corresponding chunks for a given subject as *Overlapping Image Collection (OIC)*. Look over the proposed pipeline on Figure 1, further described in the following subsections, for a clear understanding.

In this work, we develop a method that employs hierarchical Gaussian Of Gaussian (GOG) [20] and Local Maximal Occurrence (LOMO) [23] descriptors. While GOG presents a part-based model [20], [31] that subdivides a person's image into smaller segments to better build feature descriptors,

[1]We may also refer to the original full-body images, provided by the utilized datasets, as *authentic, genuine* and *primary* samples. On the contrary, images generated from overlapping blocks are commonly referred to as *derived* samples or overlapping chunks.

LOMO maximizes the occurrence of local features to obtain a consistent color representation from different camera views. The proposed approach also estimates representation features in a low-dimensional latent space using Cross-View Quadratic Discriminant Analysis (XQDA) [23], responsible for producing fast and efficient learning metrics that are also able to reduce variations in human appearance caused by different camera views. Several works in the literature assure that moving data into a common subspace is able to improving the feature discriminability and consequently the accuracy [23]–[27].

A. Training Stage

The approach loops through each individual i from the gallery set G to generate its correspondent overlapping image collection $c_i \in C$, as shown on Figure 1-A. Consequently, pairs of GOG and LOMO feature representations are obtained and merged together for each image in c_i, where each $c_i \in C$ consists of five images, particularly, the original input image and four complementary intersecting images covering mainly head, trunk, abdomen and legs. As these features are acquired for all n known subjects individually, where $n = |G|$ and $1 \leq i \leq n$, they are stored in a new analogous set of features vectors $f_i \in F$. In an equivalent manner, each f_i holds five vectors, one for each image associated with c_i. After extracting features and storing them in the feature set F, the method projects them onto a low-dimensional XQDA subspace attempting to make these feature representations more robust to the predominant person Re-ID obstacles, mainly different-camera field of views. Feature vectors from non-authentic images play an essential role on the testing stage, when machine learning algorithms are employed to improve the overall performance of the method.

B. Testing Stage

Given a probe image p, the process starts with the OIC generation. Strictly speaking, the single query image is replicated due to the addition of overlapping blocks, similarly to the training stage. The query image's data augmentation turns out to be the collection $\hat{c}$, also containing five images. Then, the proposed method extracts GOG and LOMO features for each image composing $\hat{c}$. After the concatenation of GOG and LOMO features, the approach builds a list of five features vectors, stored in a container named $\hat{f}$. In summary, the five vectors correspond to the primary full-body image and its four overlapping blocks. XQDA is then applied on the feature set $\hat{f}$ in pursuance of features comparable to the previously learned subspace. The search for the correct identity of a given probe image is organized in two stages, described as follows.

1) First step: The first step considers the initial full-body images only. It computes a similarity metric between the probe feature and all feature vectors conjoined with the gallery set's genuine samples in a k-nearest-neighbor fashion. The algorithm generates a list of candidates encompassing the distance from the current probe feature vector to every single subject enrolled in the gallery set. The list is sorted in ascending order so that the subjects that closest match the probe image come out first. Figure 1-C and Figure 3 illustrate the list of candidates for arbitrary probe samples.

2) Second step: Different from the first, the second step considers primary images and all their complementary overlapping chunks. Intersecting blocks of the original images are added in favor of increasing the number of samples available for each gallery set class. According to Figure 1-E, the algorithm picks the top two individuals of the sorted list of candidates, culminating in two OIC's from the gallery set: f_m and f_n, where m and n hold the identity of the first and second best candidates, respectively. As reported in Figure 1-F, five binary SVM classification models $m_j \in M$ are learned, each one works with a single "body part". Each feature vector from f_m is allocated to the respective SVM model as a positive class $(+1)$ whereas features from f_n are assigned to the appropriate SVM classification model as a negative class (-1). In the end, the feature vectors of the probe sample are projected onto their corresponding classification models in search of their respective response values. A majority of positive responses indicate that m is more likely to corresponds to the correct identity. Otherwise, the method swaps m and n, as demonstrated in Figure 1-G, resulting in an updated list of candidates.

C. Detailed Descriptions

In this section, we give some more details on crucial aspects of proposed approach. These elements range from overlapping image collections and similarity metrics to the list of candidates and weighted support vector machine.

1) Overlapping Image Collections (OIC): The method divides the subject image into overlapping chunks horizontally – part of the sliced image chunk reappears on the subsequent chunk. Figure 2 exemplifies the way an image is sliced in overlapping blocks. For person Re-ID tasks, it seems there is no common sense when it comes to overlapping images [32]–[34]. However, it is common to split a human full-body image among head, trunk, abdomen and legs in such arrangement where the earliest tend to be gradually more discriminative than the succeeding ones [32]. New image blocks are essential to the process of picking the most discriminative samples from a list of candidates. The mechanism of splitting a human image between overlapping chunks end up acting as a procedure to generate new samples of equivalent identity: features are extracted and processed from each image block as if they are individual images, but pertaining to the same set. Namely, they are evaluated as being part of the same class.

2) Similarity Metrics: The proposed approach starts the procedure of rearranging the candidates right after computing either the Cosine or Mahalanobis distance between the probe image and all samples of the gallery set. The Cosine Distance (CD) is calculated by the dot product defined as $cd = \vec{a}.\vec{b} = \sum_i^n a_i b_i$ therefore, it is an orientation measure and the angle $\cos\Theta$ between the vectors makes up the equation of similarity as $\cos\Theta = \frac{\vec{a}.\vec{b}}{\|\vec{a}\|\|\vec{b}\|}$. On the other hand, the Mahalanobis Distance (MD) is computed as

Fig. 2. Illustration of the generated overlapping blocks/chunks. Note that in our approach we generate new samples following a fifty-percent overlap from top to bottom only.

$md = \sqrt{(x_i - \bar{x})S_x^{-1}(x_i - \bar{x})^T}\ \forall\ x_i \in x$, where x is an one-dimensional array of size n; $\bar{x}$ is mean value for x, and S is the variance-covariance matrix. In our approach, S is obtained with XQDA's *M* parameter, which corresponds to the learned metric kernel.

3) List of Candidates: Figure 3 depicts the list of candidates for a given probe image. Each row represents a probe query where the first column holds the probe sample. Agglomerated images on the right consists of all n samples from the gallery set that closest correspond to the "unidentified" image. We can infer that the closer the gallery samples are to the left columns, the more similar they are to the probe image. Strictly speaking, one can say that the yellow-bounded probe sample (a) on Figure 3, captured with camera d_1, satisfactorily encountered its corresponding identity, taken with camera d_2. It means that the legitimate corresponding identity for the probe sample (a) is the first candidate (also delimited by yellow) among all samples from the gallery set. This is no longer the case with the probe sample (b) since its corresponding identity ended up in second place. The worst case is represented by probe sample (c) since its matching identity came to a close as the least similar sample from the gallery set, that is, the last candidate from the gallery set.

4) Support Vectors Machines (SVM): For the re-ranking step, the method learns an ensemble of five binary SVM models [6], [35], regarding top candidates only. The number five corresponds to each body part concerning the generation of overlapping chunks (dataset-provided full body, and derived images regarding head, trunk, abdomen and legs) shown on Figure 1-F and Figure 2. The proposed method assumes there are more discriminating chunks than others. With that in mind, weighted SVM models are learned and their response values are multiplied by weights $w_j \in W, 1 \leq j \leq 5$ since Yi et al. [32] claim that body chunks comprising head and trunk tend to attain better results than abdomen and legs alone, but fail to outperform the association of all five together.

IV. Experiments

In this section we carry out a thorough evaluation of our algorithm, which combines elementary distance metrics and a small ensemble of binary support vector machine models followed with majority voting to track and identify people

Fig. 3. Example of a list of candidates for three arbitrary queries: given a probe sample p (first column on the left), gallery-registered individuals $g_i \in G$ (remaining columns) are arranged in such a way that leftmost samples are more similar to p. Note that in this example, *(a)* has a Rank-1 corresponding identity, indicating that the algorithm successfully encountered the correct subject, *(b)* has its corresponding probe at Rank-2 and *(c)* is an undesirable case since it mistakes $n-1$ identities before returning the right subject at Rank-n.

across multiple security cameras. We provide a brief dataset summary in Section IV-A. Evaluation metric and Protocol are detailed in Section IV-B. In Sections IV-C and IV-D we specify Feature Descriptors and Parameters, respectively. Finally, Section IV-E reports all results, including experiments comparing our method to those from the literature.

A. Datasets

We evaluate the proposed approach on two publicly available datasets: CUHK01 and PRID450s. CUHK01 [11] was recorded in a campus environment containing around 971 individuals captured by two distinct field-of-view cameras. On the other hand, PRID450s [5] contains 450 pairs of pedestrian images. Due to the non-deterministic characteristic of the utilized protocol, the CUHK01 experiments were repeated 10 times whereas PRID450s had no less than 30 executions. The numbers of experiments vary as we set a unique upper bound runtime for both benchmarks. On account of CUHK01 having more samples and classes than PRID450s, it demands more computational time in every iteration, culminating in only 10 repetitions. Based on the majority of literature approaches comprising these datasets [20], [36], [37], dataset samples are resized to the following $height \times width$ dimensions: 160×60 for CUHK01 and 128×48 for PRID450s.

B. Evaluation Metric and Protocol

Cumulative Matching Characteristics (CMC) is the *standard* performance metric of biometric systems operating in the closed-set identification task. CMC demonstrates how often subjects show up in distinct ranks like 1, 5, 10, 50, etc. based on a matching rate. In general, X-axis indicates ranking values versus the accuracy rate depicted on the Y-axis.

TABLE I
CMC EVALUATION OF THE PROPOSED APPROACH CONSIDERING FULL BODY AND/OR DIFFERENT OVERLAPPING CHUNKS OF PRID450S DATASET.

	Chunks				**Body**				**Body+Chunks**				**Body+Chunks+SVM**			
Methods/Ranks	1	5	10	20	1	5	10	20	1	5	10	20	1	5	10	20
GOG+LOMO	72.1	89.8	94.0	97.5	72.5	90.9	95.5	98.5	73.9	90.9	94.9	98.0	75.4	90.9	95.5	98.5
GOG	68.7	87.6	93.0	97.2	66.9	87.6	93.7	97.3	70.9	89.0	93.6	97.6	71.2	87.6	93.7	97.3
LOMO	58.2	81.1	88.0	93.7	60.0	82.1	89.3	95.0	60.9	83.1	89.8	94.9	63.4	82.1	89.3	95.0

The employed protocol is similar to the one used by Paisitkriangkrai et al. [38] where all identities are randomly divided into two disjoint subsets. On CUHK01, we randomly select 486 out of 971 dataset individuals for training a new subspace. The remaining subjects' pairs of image are split between gallery and probe sets. Similarly, we pick 225 out of 450 individuals from PRID450s to compose the training set as the other subjects' image pairs end up constituting gallery and probe sets.

Table I demonstrates the proposed method's performance on PRID450s dataset considering features extracted separately. The approach's evaluation is performed taking into account each OIC (see Section III-C1). Column *Chunks* holds results for experiments discarding original samples provided by the dataset and considering overlapping blocks only. In addition, column *Body* contains the outcome of experiments comprising only authentic images supplied by the dataset. Column *Body+Chunks* encloses experiments that combine primary and derived images. More precisely, we generate overlapping image chunks to augment data and perform a simple k-nearest-neighbor algorithm in place of the SVM reranking stage. To conclude, we demonstrate the results for the approach meticulously described in Section III-B2 on column *Body+Chunks+SVM*. Note that results were gradually improved with the addition of extra data and learning models.

C. Feature Descriptors

The feature extraction representation employs both GOG and LOMO in combination with XQDA to determine subjects similarity in the new low-dimensional subspace. We come up with two approaches: one executes the feature concatenation before XQDA projection whereas the other concatenates feature vectors after. They are better described below:

- *XQDA(GOG) + XQDA(LOMO):* GOG and LOMO features vectors are extracted separately, followed by individual XQDA projections and finished up with feature vectors concatenation.
- *XQDA(GOG + LOMO):* GOG and LOMO features vectors are obtained independently, followed by their concatenation and finished up with a XQDA projection.

For each dataset evaluation outlined on Tables II and III, we only consider the best results attained with Cosine and Mahalanobis distances. There is a difference of approximately 9 p.p. and 4 p.p. at Rank-1 on CUHK01 and PRID450s datasets, respectively, in the aforementioned proposed approaches. The *XQDA(GOG + LOMO)*-based approach attained top result on CUHK01 benchmark opposing the *XQDA(GOG) + XQDA(LOMO)*-based version, which reached best result on PRID450s dataset.

The Cosine distance has shown to be more effective on the PRID450s benchmark since the combination of some factors influence the choice of distance metrics. In the case of the CUHK01 dataset, the amount of images available in the dataset can improve the correlation between samples and consequently the final XQDA projection becomes more powerful. Zhang et. al [39] state that there is a high probability that weights assigned to each dimension of the feature vector may not indicate the actual significance of that dimension. Besides, the distance measure considers the relation between each feature vector element with all other elements.

D. Parameters

This section describes the experimental configuration that differs from most relevant works in the literature. The custom settings are applied to feature descriptors, image chunks and machine learning algorithms. More details are characterized below:

1) GOG, LOMO and XQDA: We follow the same parameter criteria adopted by Matsukawa et.al [20] and Liao et.al [23] in the process of extracting GOG and LOMO feature descriptors. GOG is the general term to the fusion of different color space descriptors. XQDA has an essential parameter, a regularizer λ set as $1x10^{-5}$, that impacts how smooth and robust the subspace estimation can be.

2) OIC: The proposed approach divides the original image into four overlapping partitions. As demonstrated on Figure 2, each chunk consists of 50% of the lower part of the previous image and 50% of the upper part of the subsequent image. Figure 4 illustrates the CMC curves with the employment of the Cosine distance for separate chunks of the image. In fact, the method calculates the Cosine distance for every image part pertaining to gallery and probe sets. Based on an empirical evaluation, the results have led us to a specific chunk weight arrangement: full body, head, trunk, abdomen and legs chunks are respectively multiplied by each weight stored in the array *[0.286, 0.21, 0.168, 0.193 e 0.143]*. Some body parts, such as head and trunk, have more discriminative characteristics than abdomen and legs, justifying the reason for receiving higher weights and resulting in a bigger influence during the prediction process.

3) SVM: Multiple support vector machine models are populated with samples from the candidate list, one sample for each class where SVM classifier is trained for a specific body chunk. The SVMs are initialized with its default parameters and set with a maximum of 1×10^3 numerical optimization

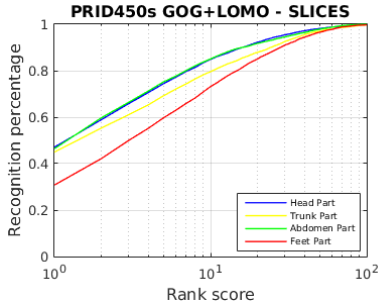

Fig. 4. CMC ranks obtained on PRID450s dataset demonstrate the degree of relevant information present in each body part on the final rank result. This chart represents the ranks using the concatenated GOG and LOMO descriptors.

iterations. The adopted *linear* kernel function ϕ computes the inner product of the transformed predictors as it can be understood as an useful way to transfer the observed features into another space. Further, the C parameter adjusts that the misclassification of training examples is set to 1.0. For large values of C, the optimization picks a smaller-margin hyperplane since it performs better at getting all the training samples classified correctly.

E. Results

We observe through the experimental results that most gallery identities that correspond to the probe query come up at the first positions in the list of candidates during the first ranking stage. This is the primary objective of the testing stage's *first step* whereas the *second step* focuses on re-ranking the list of candidates with the addition of overlapping body chunks. We contrast our method with approaches that use deep learning and handcrafted features. Table II shows the results found for the CUHK01 dataset whereas Table III outlines the outcome for PRID450s dataset. Italicized values in the tables' first rows indicate the results from deep learning approaches.

The metric used to calculate the distance between pairs of samples is directly linked to the reorganization of the candidate list. In this case, the use of classification models to estimate distances, combined with the information provided by the overlapping image collections, clearly enhances the candidate list re-ranking process, justifying our method. In some cases, where the parts of the samples are not discriminatory enough, we notice that the SVM models cannot differentiate the classes very well. In such cases, erroneous changes may occur in the reorganization of the samples in the candidate list.

The proposed approach has not outperformed previous state-of-the-art works on the CUHK01 benchmark. However, it is important to note that our approach closely matches some deep-learning-based architectures' performance [41] with a Rank-1 of 78.4% for the PRID450s dataset and a relatively good approximation for the CUHK01 dataset with 73.4% on CMC Rank-1. The method's performance neared Yang et. al [42]'s work efficiency, which uses extra information to enhance the method's overall power. We attained our best Rank-1 result of 61.1% under the $XQDA(GOG + LOMO)$ feature representation, overcoming some previous works [20], [23], [36], [38]. On PRID450s dataset, the $XQDA(GOG + LOMO)$-based approach obtained expressive results that outperformed previous literature works [20], [23], [37], [43] as it reaches a CMC Rank-1 of 75.4%. The method comprising only $XQDA(GOG)$ also attained a satisfactory result, a Rank-1 of 71.2%, but a slightly lower performance when compared to the previous algorithm. It is expected due to the fact that $XQDA(GOG + LOMO)$ combines GOG and $LOMO$ descriptors.

TABLE II
TOP RANKED APPROACHES (CMC@ RANK-R, %) ON THE SINGLE-SHOT PROTOCOL (M=1) OF CUHK01 DATASET. DEEP LEARNING METHODS ON TOP ROWS AND OUR PROPOSED ALGORITHMS AT THE BOTTOM.

CUHK01	1	5	10	20
Deep Late-fusion [40]	73.4	89.9	94.9	-
Deep L-fusion L/M/H level [41]	70.3	88.5	93.3	-
MVLDML [42]	**61.4**	**82.7**	88.9	**93.9**
GOGFusion+LDNS [36]	60.8	81.7	88.4	93.5
GOG+XQDA [20]	57.8	79.1	86.2	92.1
MetricEnsemble [38]	53.4	76.4	84.4	90.5
LOMO+XQDA [23]	49.2	75.7	84.2	90.8
OURS+XQDA(GOG+LOMO)	61.1	80.1	**90.1**	93.1
OURS+XQDA(GOG)	59.5	79.0	87.3	91.0
OURS+XQDA(LOMO)	49.0	78.8	85.1	89.2

TABLE III
TOP RANKED APPROACHES (CMC@ RANK-R, %) ON THE STANDARD PROTOCOL (P=225) OF PRID450S DATASET. DEEP LEARNING METHODS ON TOP ROWS AND OUR PROPOSED ALGORITHMS AT THE BOTTOM.

PRID450s	1	5	10	20
Deep Late-fusion [40]	78.4	94.2	96.8	-
Deep L-fusion L/M/H level [41]	77.6	93.5	96.1	-
CSPL+GOG [37]	69.2	90.4	95.5	98.2
XQDA+GOG [20]	68.4	88.8	94.5	97.8
CSPL+LOMO [37]	60.3	84.7	91.5	96.3
XQDA+LOMO [23]	59.2	83.8	90.4	95.1
Mirror-KMFA [43]	55.4	79.3	87.8	91.5
OURS+XQDA(GOG)+XQDA(LOMO)	**75.4**	**90.9**	**95.5**	**98.5**
OURS+XQDA(GOG)	71.2	87.6	93.7	97.3
OURS+XQDA(LOMO)	63.4	82.1	89.3	95.0

V. CONCLUSIONS

In our work, we use handcrafted features to treat the problem of re-identifying people. It reorganizes the ranking generated by metric distances through the addition of a small enbedding of SVM responses. The method captures what overlapping image collections have of most discriminating, repositioning potential samples from a list of candidates to the top positions of a ranks table. We used a low dimensional subspace to promote the description of the samples by means of the combination of GOG and LOMO feature descriptors. We thus evaluated our approach on a smaller data set (PRID450s) and another one of medium proportion (CUHK01), corroborating that our method brings improvements in the distinction of the samples involved and consequently in the final rank. For future work, we intend to focus on expanding our list of candidates by searching for potential samples in ranks farther from the probe.

REFERENCES

[1] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti, "Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking," *IEEE signal processing magazine*, vol. 22, no. 2, pp. 38–51, 2005.

[2] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior, and Y. Tian, "Smart surveillance: applications, technologies and implications," in *International Conference on Information, Communications and Signal Processing*, vol. 2. IEEE, 2003, pp. 1133–1138.

[3] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, 2014.

[4] Y. Yan, B. Ni, Z. Song, C. Ma, Y. Yan, and X. Yang, "Person re-identification via recurrent feature aggregation," in *European Conference on Computer Vision*. Springer, 2016, pp. 701–716.

[5] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof, "Mahalanobis distance learning for person re-identification," in *Person Re-Identification*. Springer, 2014, pp. 247–267.

[6] R. H. Vareto, S. S. Da Silva, F. D. O. Costa, and W. R. Schwartz, "Face verification based on relational disparity features and partial least squares models," in *Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference on*. IEEE, 2017, pp. 209–215.

[7] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[8] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2360–2367.

[9] B. J. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking." in *BMVC*, vol. 2, no. 5, 2010, p. 6.

[10] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification." in *Bmvc*, vol. 1, no. 2. Citeseer, 2011, p. 6.

[11] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 31–44.

[12] M. Nappi and H. Wechsler, "Robust re-identification using randomness and statistical learning: Quo vadis," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1820–1827, 2012.

[13] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3610–3617.

[14] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *European conference on computer vision*. Springer, 2014, pp. 1–16.

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

[16] 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*. Springer, 2014, pp. 536–551.

[17] B. Ma, Q. Li, and H. Chang, "Gaussian descriptor based on local features for person re-identification," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 505–518.

[18] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 144–151.

[19] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[20] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1363–1372.

[21] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*. IEEE, 2016, pp. 1249–1258.

[22] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *European Conference on Computer Vision*. Springer, 2016, pp. 791–808.

[23] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197–2206.

[24] R. Prates and W. R. Schwartz, "Kernel cross-view collaborative representation based classification for person re-identification," *arXiv preprint arXiv:1611.06969*, 2016.

[25] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*. IEEE, 2012, pp. 203–208.

[26] R. F. Prates and W. R. Schwartz, "Kernel hierarchical pca for person re-identification," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 2091–2096.

[27] R. Prates, M. Oliveira, and W. R. Schwartz, "Kernel partial least squares for person re-identification," in *Advanced Video and Signal Based Surveillance (AVSS), 2016 13th IEEE International Conference on*. IEEE, 2016, pp. 249–255.

[28] H. Iwata, K. Ebana, Y. Uga, and T. Hayashi, "Genomic prediction of biological shape: elliptic fourier analysis and kernel partial least squares (pls) regression applied to grain shape prediction in rice (oryza sativa l.)," *PloS one*, vol. 10, no. 3, p. e0120610, 2015.

[29] S. Lehtinen, J. Lees, J. Bähler, J. Shawe-Taylor, and C. Orengo, "Gene function prediction from functional association networks using kernel partial least squares regression," *PloS one*, vol. 10, no. 8, p. e0134668, 2015.

[30] S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.-N. Wang, B. Liu, and P. Phillips, "Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression," *IEEE Access*, vol. 4, pp. 7567–7576, 2016.

[31] R. Satta, "Appearance descriptors for person re-identification: a comprehensive review," *arXiv preprint arXiv:1307.5748*, 2013.

[32] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 34–39.

[33] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," *arXiv preprint arXiv:1705.04724*, 2017.

[34] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling," *arXiv preprint arXiv:1711.09349*, 2017.

[35] R. Vareto, S. Silva, F. Costa, and W. R. Schwartz, "Towards open-set face recognition using hashing functions," in *Biometrics (IJCB), 2017 IEEE International Joint Conference on*. IEEE, 2017, pp. 634–641.

[36] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1239–1248.

[37] J. Dai, Y. Zhang, H. Lu, and H. Wang, "Cross-view semantic projection learning for person re-identification," *Pattern Recognition*, vol. 75, pp. 63–76, 2018.

[38] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1846–1855.

[39] D. Zhang and G. Lu, "Evaluation of similarity measurement for image retrieval," IEEE, pp. 928–931, 2003.

[40] A. R. Lejbølle, K. Nasrollahi, and T. B. Moeslund, "Late fusion in part-based person re-identification," in *Proceedings of the 9th International Conference on Machine Learning and Computing*. ACM, 2017, pp. 385–393.

[41] ——, "Enhancing person re-identification by late fusion of low-, mid- and high-level features," *Iet Biometrics*, 2017.

[42] X. Yang, M. Wang, and D. Tao, "Person re-identification with metric learning using privileged information," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 791–805, 2018.

[43] Y.-C. Chen, W.-S. Zheng, and J. Lai, "Mirror representation for modeling view-specific transform in person re-identification." in *IJCAI*, 2015, pp. 3402–3408.