A Neuromorphic Person Re-Identification Framework for Video Surveillance

Aparajita Nanda[†], Pankaj K. Sa[‡], Suman K. Choudhury[†], Sambit Bakshi[‡], and Banshidhar Majhi[‡]

[†‡]Computer Science and Engineering, National Institute of Technology, Rourkela – 769008, India
[†]{aparajita1.nanda, sumanchoudhury.nitr}@gmail.com
[‡]{PankajKSa, BakshiSambit, BMajhi}@nitrkl.ac.in

September 6, 2017

Abstract

In this article, we present a Neuromorphic Person Re-Identification (NPReId) framework to establish the correspondence among individuals observed across two disjoint camera views. The framework comprises three modules (*observation*, *cognition*, and *contemplation*), inspired by the Form-and-Color-and-Depth (FACADE) theory model of object recognition system. In *observation* module, a semantic partitioning scheme is introduced to segment a pedestrian into several logical parts, and an exhaustive set of experiments have been carried out to select the best possible complementary feature cues. In *cognition* module, an unsupervised procedure is suggested to partition the gallery candidates into multiple consensus clusters with high intra-cluster and low inter-cluster similarity. A supervised classifier is then deployed to learn the relationship between each gallery candidate and its associated cluster, which is subsequently used to identify a set of inlier consensus clusters. This module also includes weighing of contribution of each feature channel towards defining a consensus cluster. Finally, in *contemplation* module, the contributory weights are employed in a correlation-based similarity measure to find the corresponding match within the inlier set. The proposed framework is compared with several state-of-the-art methods on three challenging datasets: VIPeR, iLIDS-VID, and CUHK01. The tabular results alongside the performance curves demonstrate the superiority of NPReId over the counterparts.

Index Terms: Surveillance; Person re-identification; Recognition; Consensus clustering; Similarity measure; Feature extraction; Information gain; CMC.

1 Introduction

In the last two decades, there has been a tremendous growth in the use of visual surveillance systems. The research community in academia, as well as R&D organizations, are actively involved in making the video surveillance automated and intelligent. Object detection, tracking, recognizing objects of interest, understanding and analyzing their activities are some of the key ingredients of a smart surveillance system. With the advent of the multi-camera networks, newer issues have surfaced that demand deeper understanding and significant research. Person re-identification is one such issue, which is about re-identifying a previously observed person that leaves the field of view (FoV) of one camera and enters the FoV of another camera, or re-enters the FoV of the same camera after a period of time. In particular, a given probe image is searched in the set of available gallery images, and the least distant image (with maximum similarity) becomes the potential match; an abstraction of the same is depicted in Figure 1. Use of biometric traits such as face, gait, periocular, fingerprint, *etc.* seem to be good candidates to solve the person re-identification problem. However, surveillance systems do not have the luxury of constrained environment where the images could be recorded as desired. The images are usually of very poor resolution that limits the efficacy of biometric systems considerably. In addition, person re-identification suffers from few severe challenges, such as, indistinguishable attire, unalike appearance, pose variations, varying background, partial occlusion *etc*; images in Figure 2 show some of the typical scenarios where the above challenges can be well observed.

Quite a significant amount of work on person re-identification have been reported in the recent pasts. Prosser *et al.* formulate a bidirectional brightness transfer function to compute a chromatic based mapping across the disjoint camera views [1]. Farenzena *et al.* segment a human silhouette with horizontal-vertical symmetry followed by chromatic feature extraction from each segmented body part [2]. Cheng *et al.*, in their work, apply the pictorial structure to locate different body parts [3]. Kviatkovsky *et al.* propose an invariant color signature in the log-chromaticity space by considering the color distribution under different lighting conditions [4]. Ma *et al.* suggest a biological covariance (BiCov) descriptor to

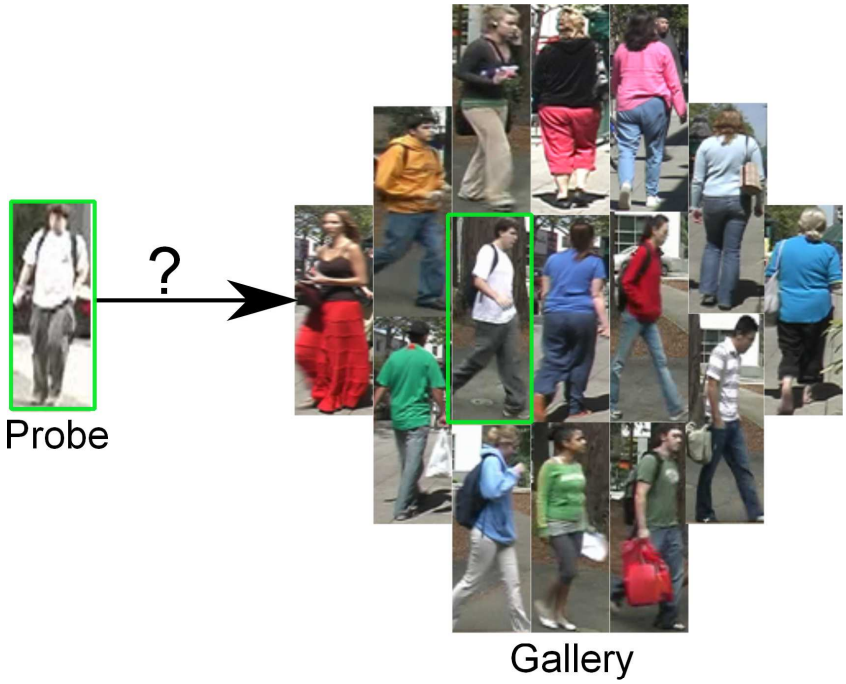

Figure 1: An abstraction of person re-identification — A probe is compared with all gallery candidates to find the exact match.

address the problem with illumination change [5]. Shi *et al.* formulate a multi-level adaptive correspondence method for handling the misalignment of body parts [6]. In another work, a part based segmentation approach is suggested to solve the problem with pose misalignments [7]. Liao *et al.* model a stable feature representation based on the idea of maximizing the horizontal occurrence of local features to counter the problems with varying viewpoints across the camera views [8]. Zheng *et al.* [9] introduce a probabilistic relative distance comparison (PRDC) model to formulate the re-identification task as a distance learning problem. Koestinger *et al.* model the re-identification problem as a task of metric learning with equivalence constraints [10]. Li *et al.* model a quadratic decision function for metric learning [11]. In another work, a kernel-based distance learning approach is presented to improve the re-identification accuracy [12]. Subsequently, the discriminative and representative patches are collected for feature learning [13]. Zaho *et al.* propose an unsupervised salience learning model to learn the salience regions of human appearance [14], [15]. Wang *et al.* design a video ranking model by simultaneously selecting and matching the reliable space-time features from the image sequence [16]. Zheng *et al.* propose a score-level fusion scheme that automatically selects an appropriate set of features from the unlabeled data [17]. An *et al.* formulate a robust canonical correlation analysis to map the samples from two disjoint views into a subspace followed by similarity matching [18], [19].

Most of the existing literature focus on the following two areas — (i) a robust pedestrian signature in terms of invariant feature representation, and (ii) an efficient similarity measure to find the potential match. It has been observed that pedestrian signature with single feature is not adequate to counter the challenges posed by person re-identification. Therefore, existing models prefer the use of a set of features to strengthen the ability of the signature. The set should be so chosen that the individual features complement each other and also enough care should be given so that the combination do not lead to overfitting. The second issue under focus is on the similarity measure, where exhaustive search of probe with all the gallery candidates seems to be the most intuitive approach. However, this process is not only time consuming but also tend to produce inaccurate match owing to the feature space limitations. Our proposition, in this article, thrusts upon — (i) selection of complementary features combination, (ii) looking for an inlier subset of gallery candidates for a given probe, where the probability of finding the match is very high.

FACADE (Form-And-Color-And-Depth) [20], a neural network theory, supports the biological cogency of an object-based model, and presents a framework that combines the observation of visual perception with the recognition system. The left hand side diagram of Figure 3 illustrates the FACADE theory with three modules. Boundary contour system (BCS) performs the segmentation of foreground from its underlying background in the visual cortex. Feature contour system (FCS), on the other hand, extracts the feature details of the object boundary in terms of color and orientation. The object recognition system (ORS), based on the adaptive resonance theory, reinforces both BCS and FCS on the correct recognition of the object. In this paper, we present a neuromorphic person re-identification

Figure 2: Different appearances of the same person captured from two disjoint camera views. Images in a specific column denote the same person. (a) images with similar color attire, (b) images with variation in appearances, (c) image-pairs with variations in pose and viewpoint, (d) image-pairs with illumination and background variations, and (e) partially occluded image-pairs.

system (NPReId) following the FACADE theory that comprises three interactive modules – *observation*, *cognition*, and *contemplation* as shown in the right hand side of Figure 3. The observation module suppresses the background and extracts the chromatic and texture details from the segmented pedestrian. The cognition module projects the psychological result of observation to learn the underlying pedestrian signature. The results of observation and cognition modules are forwarded to the contemplation module that recognizes the correct match for any individual.

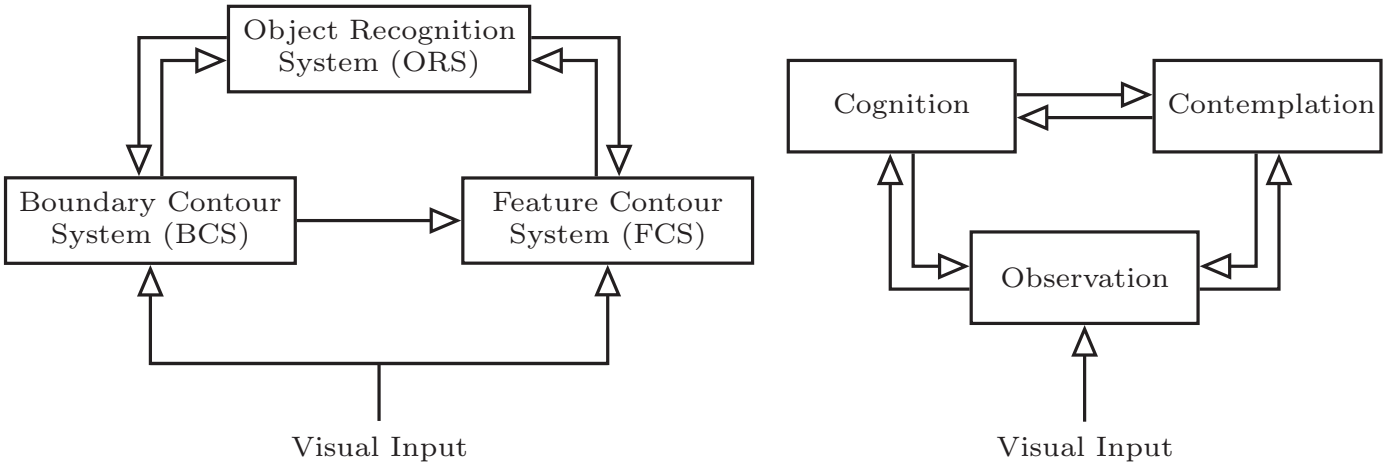

Figure 3: The FACADE modules. Left: the BCS-FCS includes the background segmentation, foreground feature extraction, which are further recognized by the ORS. Right: the interactive modules of proposed re-identification system (NPReId) based on FACADE theory.

The rest of this article is organized as follows. The proposed NPReId system is elaborated in Section 2. Simulation results on standard datasets along with other state-of-the-art methods are presented in Section 3. Finally, concluding remarks are given in Section 4.

2 The Proposed NPReId Framework

In this paper, we present a NPReId framework, in line with the FACADE theory, to establish the correspondence between a probe and a subset of gallery images. Figure 4 depicts the overview of the proposed framework. In *observation* module, we introduce a part-based segmentation, where the entire body is semantically partitioned into seven segments. A comparative analysis has been carried out to select an appropriate feature set to represent a pedestrian signature. Then, in *cognition* module the gallery set is partitioned into a number of consensus clusters following the K-means method and a cluster ensemble approach. The principle of information gain is suitably formulated to compute the contribution of each feature channel towards defining its associated cluster. The relationship of each gallery feature vector with the

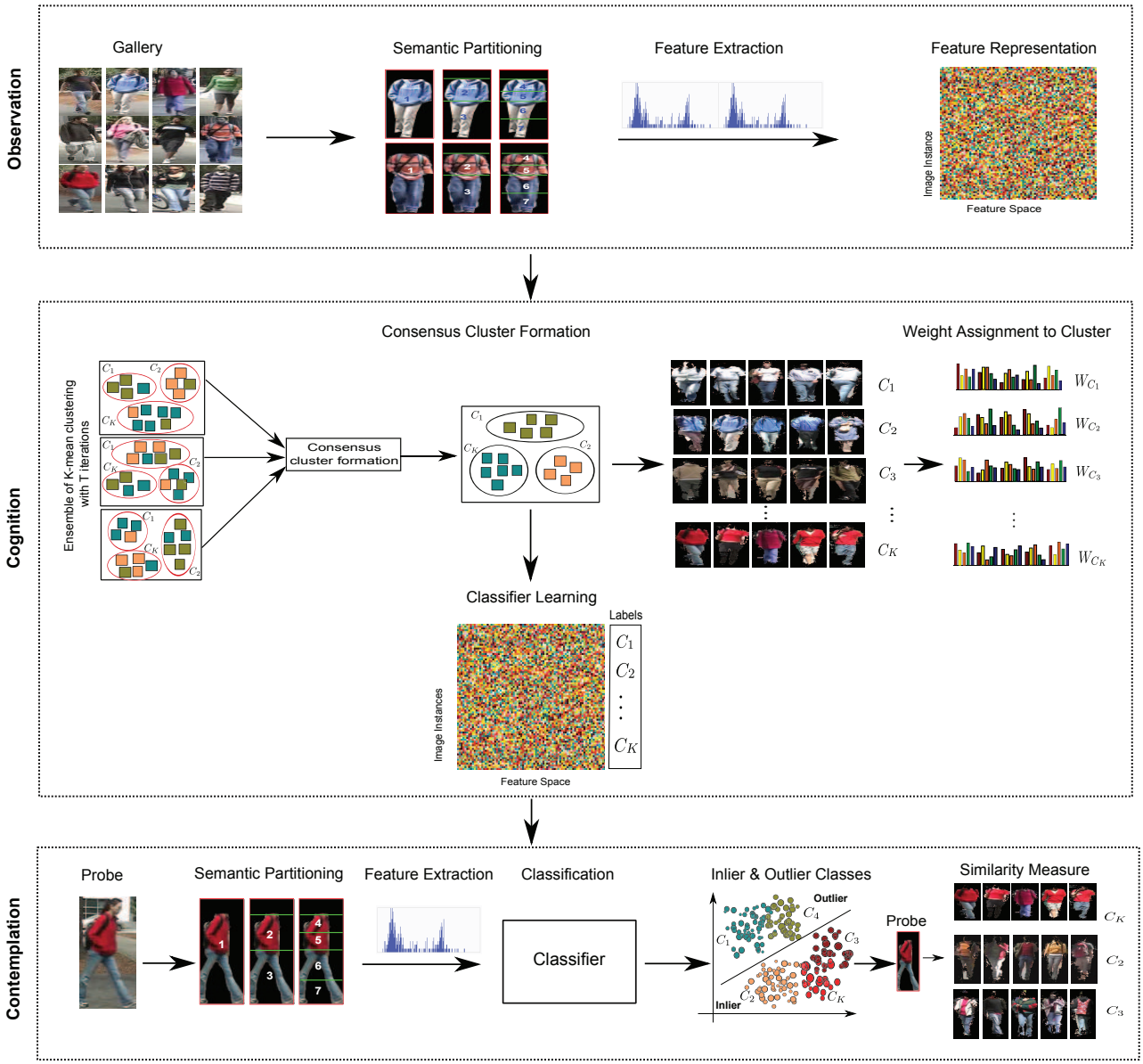

Figure 4: Overview of the proposed NPReId framework with three interactive modules — *Observation*, *Cognition*, and *Contemplation* along with the steps in each module.

corresponding cluster is learned using a classification model. During the *contemplation* stage, the learned model selects a set of inlier clusters for a given probe. A correlation based similarity measure is then applied to find the exact match within the filtered set.

2.1 Observation

The observation module includes some preprocessing tasks like background suppression, semantic partitioning, followed by feature representation.

2.1.1 Semantic partitioning of body structure

The cluttered background around a pedestrian image within the bounding box often leads to an erroneous feature representation. Therefore, we apply the STEL generative model [21], a preprocessing operation, to suppress the background content prior to semantic partitioning.

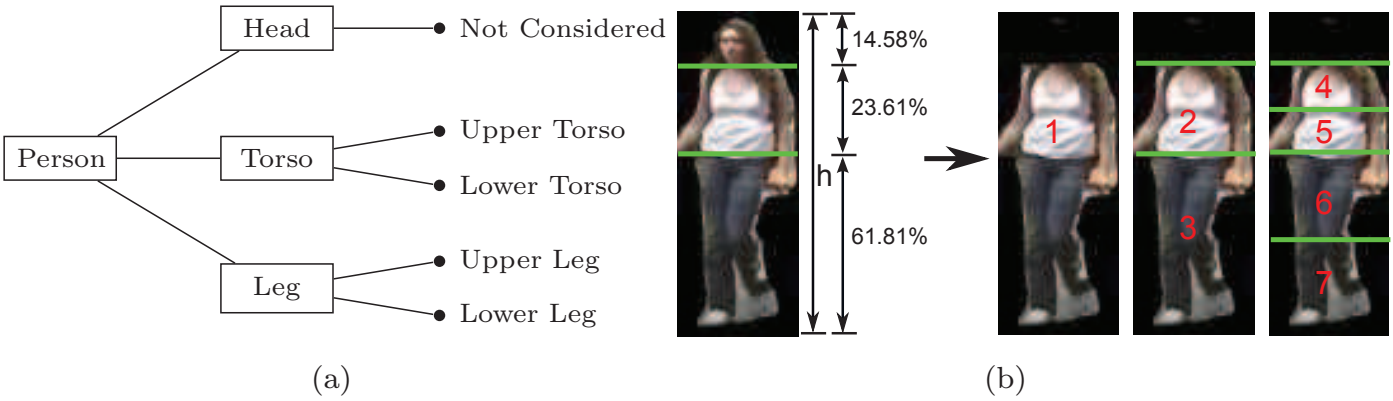

Figure 5: Semantic partitioning of the body structure

A holistic feature representation often leads to false match in case of partial occlusion. Moreover, various clothing fashion together with numerous pose yield a number of possible instances in pedestrian appearance. Therefore, the entire body needs to be semantically partitioned into various local segments prior to feature extraction.

In our work, we follow the Golden ratio (1.6180339887) principle of human body that partitions the entire body into three semantic segments: the head, the torso, and the leg at 14.58%, 23.61%, and 61.81% of the total height of a pedestrian. Person re-identification primarily relies on the appearance cues (attire similarity), and thereby we exclude the head portion that lacks any information because of poor resolution. Both torso and leg portions are encoded together as well as individually to take the advantages of holistic and part based representation. The torso and leg portions are further subdivided into two equal sized horizontal strips to extract information at a finer level. In this way, we partition a human body into seven logical segments, need to be encoded during feature representation, as shown in Figure 5.

2.1.2 Feature extraction

In person re-identification, complementary appearance cues need to be integrated to generate a robust feature representation. Usually, the invariant chromatic details along with the texture patterns are incorporated for feature encoding.

We consider multiple feature channels across the seven semantic segments, as discussed earlier, to create a robust feature signature for a pedestrian. The color channel includes Hue-weighted-Saturation[1] (HwS) [22] and CbCr, the intensity channel includes Y, and the texture channel, adapted from [23], comprises a set of eight Gabor and thirteen Schmid filters. RGB has also been considered for both color and intensity representation. All the feature channels, stated above, are quantized into 16-bin histogram. We conducted two experiments with the hypothesis that the most suitable features combination would produce highest result under any mediocre distance measure; accordingly, we choose the L_1-norm as the similarity metric. The datasets VIPeR [24], iLIDS-VID [16], and CUHK01 [25] are taken into consideration for this experimental purpose. The details of these datasets are given in Section 3.1.

In the first experiment, the performance of individual feature channel, for VIPeR dataset, is compared in terms of cumulative matching characteristics (CMC) curve as shown in Figure 6. It can be observed that HwS produces superior result over its counterparts with an early convergence at rank 180. The performances of CbCr and texture channels are also comparable to HwS. However, both RGB and Y channels fail to yield satisfactory result; the failure of which may be attributed to the intensity based features that suffers from the problem of shadow and light illumination change. The second experiment combines the above channels, for the same VIPeR dataset, to find the best possible features combination as shown in Figure 6. The other two datasets also result in similar observations as shown in Figures 6 and 6.

It can be seen that HwS + CbCr + Texture produces better result in comparison to other combinations. Therefore, we consider the 24 feature channels (1 HwS + 1 Cb + 1 Cr + 8 Gabor + 13 Schmid) across the seven semantic segments. More precisely, each pedestrian image is represented with d-dimensional feature vector, where $d = f \times b$, f denotes the number of feature channels and b denotes the dimension of each channel. In our case, $f = 24$ channels $\times$ 7 segments $=$ 168 , $b = 16$, and $d = 2688$.

2.2 Cognition

The cognition module includes consensus cluster formation, classifier learning, and weight assignment.

[1]Hue-weighted-Saturation: Hue histogram where each hue sample is weighted by its corresponding saturation value.

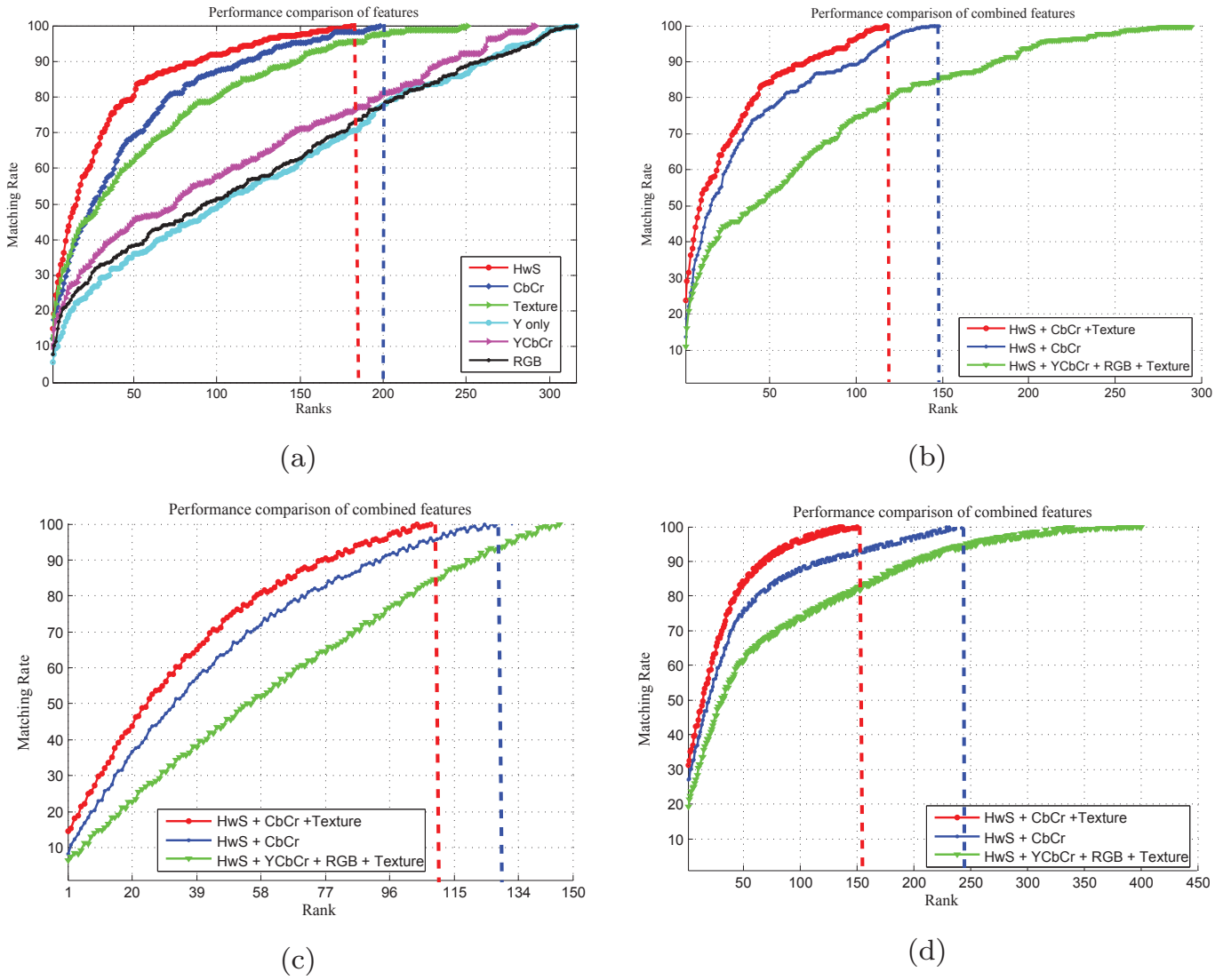

Figure 6: Comparative analysis of CMC curve: (a) across each individual feature channel for VIPeR, (b) across various features combination for VIPeR, (c) across various features combination for iLIDS-VID, (d) across various features combination for CUHK01.

2.2.1 Consensus clusters formation followed by classifier learning

In this section, we first apply an unsupervised procedure to partition the gallery set into a number of consensus clusters with high intra-cluster similarity and high inter-cluster deviation, where each cluster comprises a subset of look-alike gallery candidates having similar attributes. A classifier is then employed to learn the supervised relationship between each gallery image and its associated cluster. The procedure of consensus cluster formation and classifier learning is enumerated below.

(a) We first apply the K-means clustering to partition the gallery set into K disjoint sets. It can be realized that the choice of initial cluster centers, in K-means, greatly influences the resultant clusters. This issue is alleviated following the principle of Central Limit Theorem (CLT), wherein the K-means clustering is performed sufficiently large number of times (say T) with random initialization of cluster centers. This operation yields a total of $T \times K$ clusters.

(b) We then apply the consensus based meta-graph clustering algorithm (MCLA) [26], an approach to re-cluster the clusters, to merge the $T \times K$ clusters into K consensus clusters $\{C_1, C_2, \cdots, C_K\}$.

(c) A classifier model is then built using support vector machine (SVM) of Gaussian kernel to learn the relationship between the gallery feature vectors $I = \{I_1, I_2, \cdots, I_N\}$ and their corresponding cluster labels $C = \{C_1, C_2, \cdots, C_K\}$. This learned model has latter been used in the contemplation stage of the framework.

2.2.2 Weight assignment to each of the consensus cluster

This section analyses each feature channel relative to their contribution towards defining a cluster. We first apply information gain principle to quantify each feature bin and subsequently modify with suitable aggregation for each feature channel.

Let $\phi = (I, O)$ denotes the training pair with I as the gallery feature set ($I = \{I_1, I_2, \cdots, I_N\}$) and O as the corresponding cluster labels. The label O_i for a consensus cluster C_j ($j = 1, 2, \cdots, K$) is made in congruent with the following equation —

$$O_i = \begin{cases} 1 & \text{if } I_i \in C_j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Each feature vector I_i is represented with d-attributes ($I_i = \{I_i^\alpha\}$, $\alpha = 1, 2, \cdots, d$). According to the principle of information gain, the contribution of an attribute α with respect to a training pair ϕ can be expressed as —

$$l(\phi, \alpha) = H(\phi) - \sum_{e \in E(\alpha)} \left(\frac{|\{I_i \in \phi | I_i^\alpha = e\}|}{|\phi|} \cdot H\left(|\{I_i \in \phi | I_i^\alpha = e\}|\right) \right) \tag{2}$$

where $H(\phi)$ measures the entropy of the training set; $E(\alpha)$ denotes the set of all possible values of an attribute α. The above expression yields a d-dimensional vector $L = \{l_1, l_2, \cdots, l_d\}$ that signifies the relative contribution of each feature attribute towards defining a consensus cluster. It can be observed that larger-contributory attributes highlight the commonality among the images within a consensus cluster. In other words, the low contributory attributes are more informative in distinguishing the images within the same cluster. Accordingly, while searching a probe within the look-alikes of a consensus cluster, assignment of higher priority to attributes with low contribution becomes an obvious choice. Hence the above vector L is complimented to represent the required weight vector $W = \{w_1, w_2, \cdots, w_d\}$ as —

$$W = 1 - L \Big/ \sum_{i=1}^{d} l_i \tag{3}$$

The dense weight vector W of d-dimension needs to be further modified to quantify the contribution across feature channels with f-dimension. This dense to coarse transformation is essential as the similarity measure, employed in the Contemplation module, takes the cumulative result of correspondence across each feature channel. Accordingly, a modified weight vector $Q = \{q_1, q_2, \cdots, q_f\}$ of f-dimension is prepared by using the disjunctive aggregation (maximum) over each of the corresponding b weights of the existing vector W. Mathematically,

$$q_i = \max(\{w_t\}) \tag{4}$$

where $t = \{((i-1) \times b) + 1, ((i-1) \times b) + 2, \cdots, ((i-1) \times b) + b\}$. Use of max operation to evaluate the disjunction of properties means that the grading of the most satisfied property will reflect the global level of satisfaction.

2.3 Contemplation

In this module, the learned classification model, as discussed in Section 2.2.1, is employed to find a set of inlier consensus clusters for a given probe. Then, a correlation based weighted similarity measure is applied to find the exact match within the set of inlier clusters. The subsequent paragraphs detail both steps in sequel.

A probe feature vector is first subjected to the learned model that assigns a classification score to each of the K consensus clusters; the probability of belongingness becomes higher as the classification score increases. Accordingly, the probe is associated with the closest consensus cluster that yields the maximum score. However, the learned model may not be 100% accurate. It may so happen that the desired gallery image may be available in another cluster which may not yield the maximum score, however, comparable to it. Therefore, we need to select a set of inlier consensus clusters with high classification scores, rather than only the closest one, where the probability of finding the match is very high. We suggest an algorithm, based on the application of Z-Score labeling, to find the set of inlier clusters for a given probe, enumerated in Algorithm 1.

The last step of our framework compares a given probe within the set of inlier clusters to find the best possible match. We adapt the Quadratic-Chi histogram distance measure (χ_{quad}) [27] where the correlation of relative bin distribution of a feature channel along with the bin-wise similarity is taken into consideration. In addition, the contribution of each feature channel, in terms of weight, is incorporated in χ_{quad} to strengthen its ability in distinguishing look-alike gallery candidates within the inlier clusters. Mathematically, the distance between a probe feature vector $I_p = \{U_1, U_2, \cdots, U_f\}$ and gallery feature vector $I_g = \{V_1, V_2, \cdots, V_f\}$ is given by —

$$\mathcal{D}(I_p, I_g) = \sum_{i=1}^{f} q_i \cdot \chi_{quad}(U_i, V_i) \tag{5}$$

Algorithm 1: Computation of inlier clusters

input : Set of K consensus clusters: $\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$,
A probe feature vector I_p, and Classification model M.
output : A set of inlier consensus clusters $\tilde{\mathcal{C}}$, where $\tilde{\mathcal{C}} \subset \mathcal{C}$

$\tilde{\mathcal{C}} \leftarrow \phi$; `// Initialize` $\tilde{\mathcal{C}}$ `with empty set`
Apply probe feature vector I_p on classification model M that results in K classification scores $S = \{s_1, s_2, \cdots, s_K\}$ with respect to the set of consensus clusters $\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$;
$s_{\max} \leftarrow \max(S)$; `// Extract the maximum score from` S
$\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} \cup \{C_j\}$, where $s_j == s_{\max}$;
`// include the cluster with maximum classification score in` $\tilde{\mathcal{C}}$
$S \leftarrow S - \{s_{\max}\}$; `// Exclude` $s_{\max}$ `from` S
Create a vector A to store m random numbers ($m \geq 10$) following a normal distribution with mean $\mu = C_j$ and standard deviation $\sigma = 1$;
`//` A `is a vector with no outlier samples`
for $i \leftarrow 2$ *to* K, **do**
  $s_{\max} \leftarrow \max(S)$; `// Extract the next maximum score from` S
  $Z_{\mu,\sigma} \leftarrow (s_{\max} - \mu)/\sigma$;
  if $|Z_{\mu,\sigma}| \leq 2.5$ **then**
    `// an empirical threshold of belongingness in` Z`-Score labeling`
    $\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} \cup \{C_j\}$, where $s_j == s_{\max}$;
    $A \leftarrow A \cup \{s_{\max}\}$; `// include` $s_{\max}$ `in` A
    $S \leftarrow S - \{s_{\max}\}$;
  else
    break;

The probe that has the least distance $\mathcal{D}$ in the inlier gallery set is considered as the corresponding match.

3 Experiments and Analysis

The effectiveness of neuromorphic person re-identification framework (NPReId) is validated through an exhaustive set of experiments on three standard datasets. The results are also compared with some of the state-of-the-art methods. We also analyze the cases where our method does not produce satisfactory results. Prior to all these, we present a brief overview on the datasets used in the experiments.

3.1 Datasets and state-of-the-art methods

Three benchmark datasets namely, *Viewpoint Invariant Pedestrian Recognition* (VIPeR, [24]), *iLIDS Video re-IDentification Dataset* (iLIDS-VID, [16]), and *Campus dataset* (CUHK01, [25]), are used for the experimental evaluation. The details of the datasets, including the number of images and the kind of challenges they pose, are enumerated in Table 1. In addition, we compare the proposed NPReId framework with different sets of existing methods across three different datasets; for each dataset, we select few state-of-the-art methods where the respective articles implement the underlying dataset. The methods that we select for VIPeR dataset include LOMO+XQDA [8], MLACM [6], eBiCov [5], CLSVM [7], MidLevel [13], LADF [11], ColorInv [4], SalMatch [15], Salience [14], KISSME [10], PCCA [28], PRDC [9], CPS [3], SDALF [2], ELF [23], and PRSVM [29]. We have chosen the following methods for the iLIDS-VID dataset: the supervised approach (MidLevel [13], PRDC [9]), unsupervised approach (Salience [14], SalMatch [15], CPS [3])and multi-shot approaches (MS-SDALF [2], MS-Color+RSVM [16], MS-Color+LBP+RSVM [16]). Similarly, the simulated methods chosen for CUHK01 dataset are: Semantic [17], ROCCA [18], PRRD [19], KML [12], FUSIA [30], MidLevel [13], SalMatch [15], Salience [14], KISSME [10], PRDC [9], CPS [3], SDALF [2], LMNN [31], and ITML [32].

The earmarked gallery images of each dataset is first split into K number of clusters using the K-means algorithm; the biasness with the choice of cluster centers are alleviated by clustering sufficiently large number of times (T=200) with random initialization of cluster centers. Then, the meta-graph clustering algorithm (MCLA) is applied to re-cluster the $T \times K$ clusters to a set of K consensus clusters. We experimentally set $K = 10, 12, 13$ for the VIPeR, iLIDS-VID, and CUHK-01 datasets respectively. Few samples of consensus clusters across the three datasets are reflected in Figure 7, 8, and 9; the appearance similarity among the members of each consensus cluster is very much apparent in these figures.

Table 1: Simulated datasets including the number of images and underlying challenges

Name	# images	Resolution	Challenges
VIPeR, [24]	1264 (632 image pairs)	128 × 48	Viewpoint variation Illumination change
iLIDS-VID, [16]	600 (300 image pairs)	128 × 64	Illumination change Similarity in clothing cluttered background Partial occlusion
CUHK01, [25]	1942 (971 image pairs)	160 × 60	Pose variations Illumination change

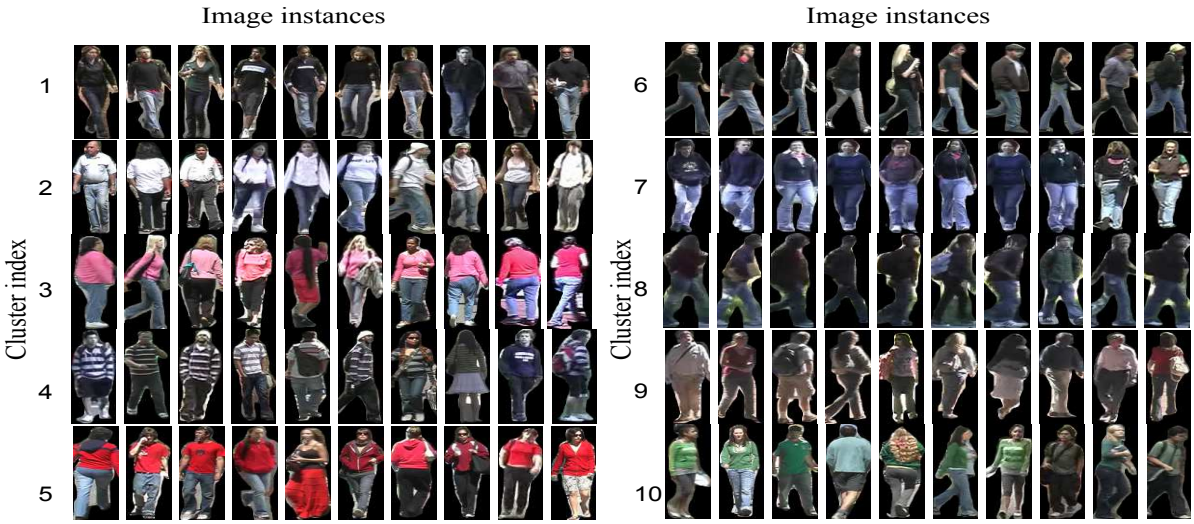

Figure 7: Sample consensus clusters in VIPeR dataset with unlabeled pedestrians

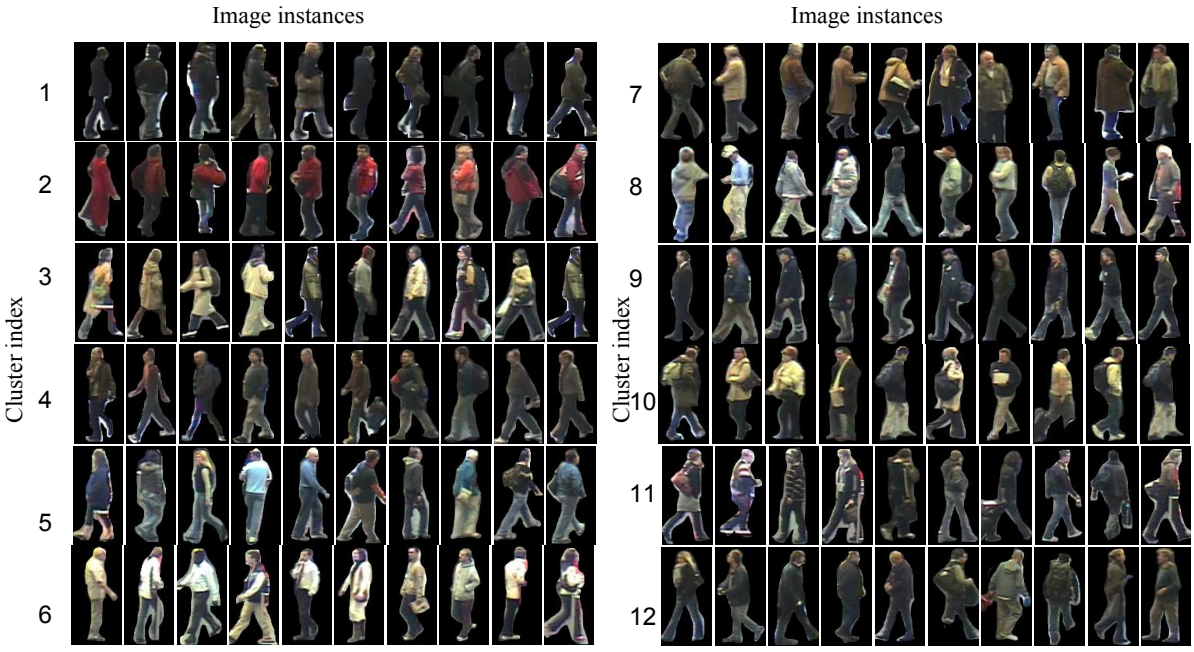

Figure 8: Sample consensus clusters in iLIDS-VID dataset with unlabeled pedestrians

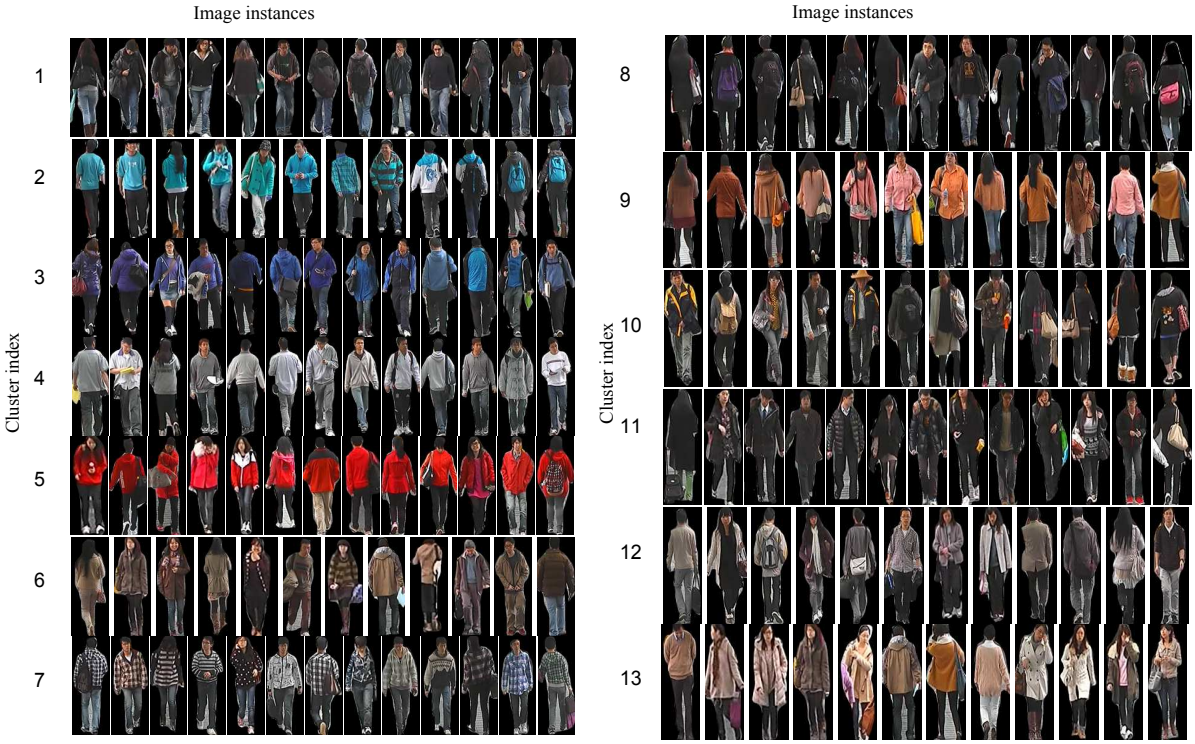

Figure 9: Sample consensus clusters in CUHK01 dataset with unlabeled pedestrians

3.2 Results analysis

Our experiments on the three datasets follow the evaluation protocol in [23]. We partition the dataset into two even parts: 50% as the gallery set for training and 50% as the probe set for testing. We conduct a set of ten trials to create ten different gallery sets and probe sets. In each trail, One of each image pair is randomly picked into the gallery set G and the other to the probe set P. The average recognition rate over these ten trials is justified through the cumulative matching characteristics (CMC) curve. CMC plots the recognition rate versus the rank; for example: Rank-r recognition rate signifies the cumulative expectation of recognition rate of all ranks upto r. The CMC curves for VIPeR, iLIDS-VID, and CUHK01 are plotted in Figures 10, 11, and 12 respectively. The tabular results of the proposed NPReId framework along with the existing schemes are compared in Tables 2, 3, and 4.

Table 2: Recognition rates (%) on the VIPeR dataset with 316 image-pairs

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$
NPReId	43.36	72.12	85.21	94.05
LOMO+XQDA [8]	40.00	68.90	81.50	91.10
MLACM [6]	34.87	59.27	70.19	81.77
eBiCov [5]	24.34	46.75	58.48	71.17
CLSVM [7]	17.09	38.61	52.53	68.35
MidLevel [13]	29.10	52.30	65.90	79.90
LADF [11]	29.30	61.00	76.00	88.10
ColorInv [4]	23.51	43.04	55.16	69.59
Salience [14]	26.31	46.60	58.86	72.77
SalMatch [15]	30.16	52.31	65.54	79.15
PCCA [28]	19.27	47.00	65.00	79.00
KISSME [10]	19.60	47.00	62.60	78.00
PRDC [9]	15.66	38.42	53.86	70.09
CPS [3]	21.84	44.00	57.21	71.00
SDALF [2]	19.87	38.90	49.37	65.73

It can be observed that the proposed NPReId framework, at rank-01 possesses atleast 43% recognition rate in case of VIPeR and CUHK01 dataset, however limits to 35% only in iLIDS-VID dataset. The reduced performance rate is attributed to more challenges in the latter case.

We further conduct a failure analysis to enumerate potential causes of false match.

(i) There could be some scenarios where human intelligence even fail to recognize a matched pair. An illustration of such instances are well depicted in Figure 13. It may be observed that the appearance of an individual across the disjoint camera views look completely different due to acute variation in pose and viewpoints.

Table 3: Recognition rates (%) on the iLIDS-VID dataset with 150 image-pairs

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$
NPReId	35.87	46.52	59.50	72.38
MS-color+RSVM [16]	16.40	37.30	48.50	62.60
MS-color+LBP+RSVM [16]	20.00	44.00	52.70	68.00
MS-SDALF [2]	5.10	19.00	27.10	37.90
MidLevel [13]	11.70	29.00	40.30	53.40
SalMatch [15]	8.01	24.80	35.40	52.90
Salience [14]	10.12	24.82	35.45	52.92
PRDC [9]	7.44	16.21	23.44	34.17
CPS [3]	7.32	15.31	21.52	30.20

(ii) Cluttered background in the bounding box of a pedestrian may lead to over-fitting. Therefore, a segmentation task is often preferred to suppress the background content prior to feature encoding. In some extreme cases, where the background and foreground are scarcely differentiable, the segmentation algorithm fails to extract the pedestrian neatly as shown in Figure 13.

Our future work concentrates on addressing the above challenges. The issues of pose and viewpoint variation are inherent to the single-shot domain. This can possibly be alleviated in the multi-shot environment, where the availability of multiple images of each individual shall lead to a robust feature representation. Effectiveness of background suppression is often limited by the poor resolution of the still images. Exploiting the motion cues in video frames may lead to better segmentation of pedestrian images and thereby reducing the false match.

4 Conclusion

In this article, we present a neuromorphic framework, inspired by FACADE theory, to re-identify persons across disjoint camera views. Our contribution concentrates on two major aspects — (i) discovering a set of complimentary cues that strengthen the resulting feature descriptor, (ii) recovering a subset of gallery candidates with high probability of retrieving the corresponding match. The proposed NPReID framework operates the above steps in sequel. The Golden ratio principle of human analogy is applied to counter the problem with pose variation and partial occlusion, where a pedestrian is partitioned into seven logical segments in a coarse to fine-manner. The efficacy of various feature channels are first analyzed individually, and subsequently a complementary features combination is decided through an exhaustive simulation across three benchmark datasets. An unsupervised procedure is suggested to partition the large gallery set into a number of consensus clusters with high intra-cluster and low inter-cluster similarity. A classifier is then employed to learn the association between each gallery feature vector and its corresponding consensus cluster. The

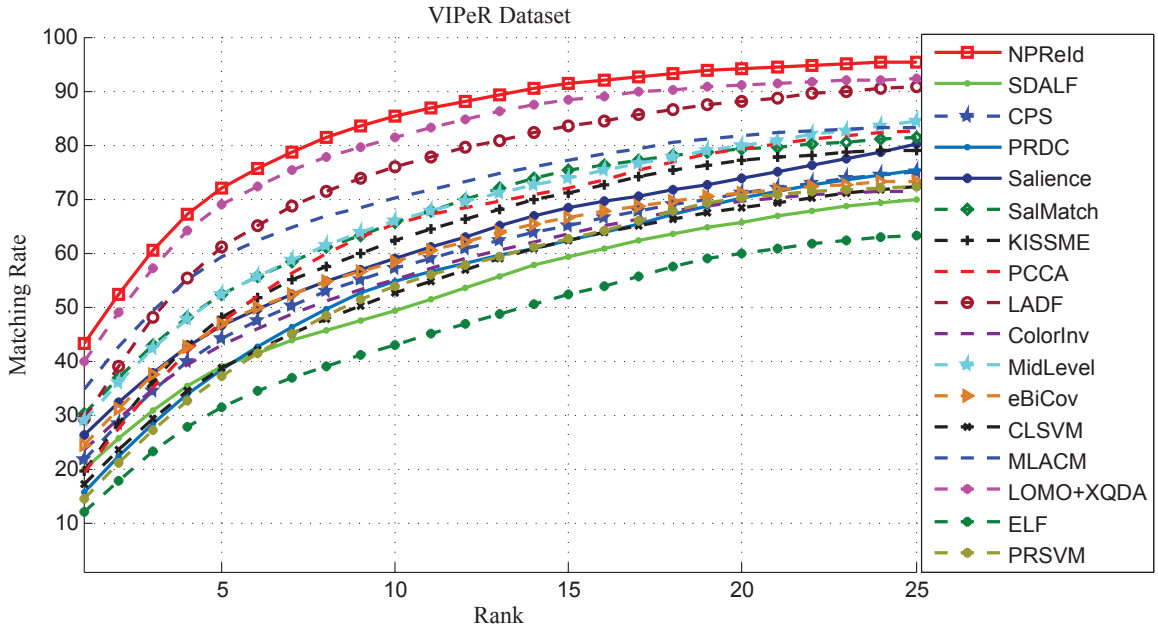

Figure 10: CMC curves for VIPeR dataset

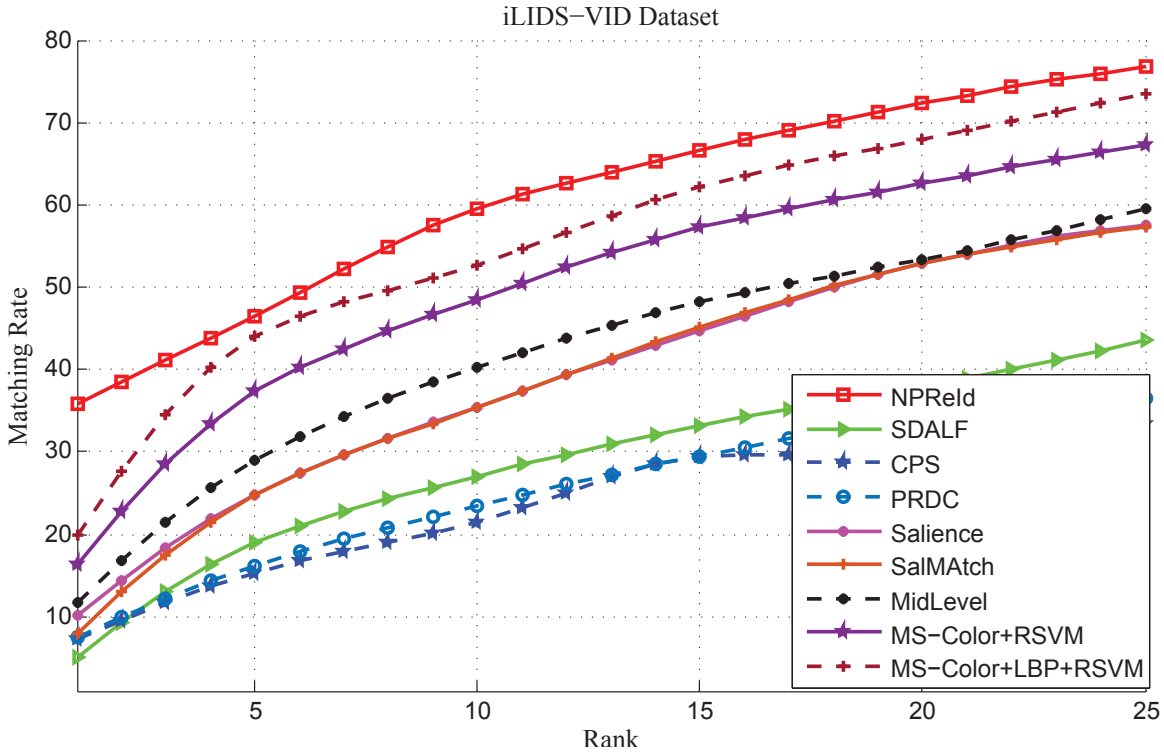

Figure 11: CMC curves for iLIDS-VID dataset

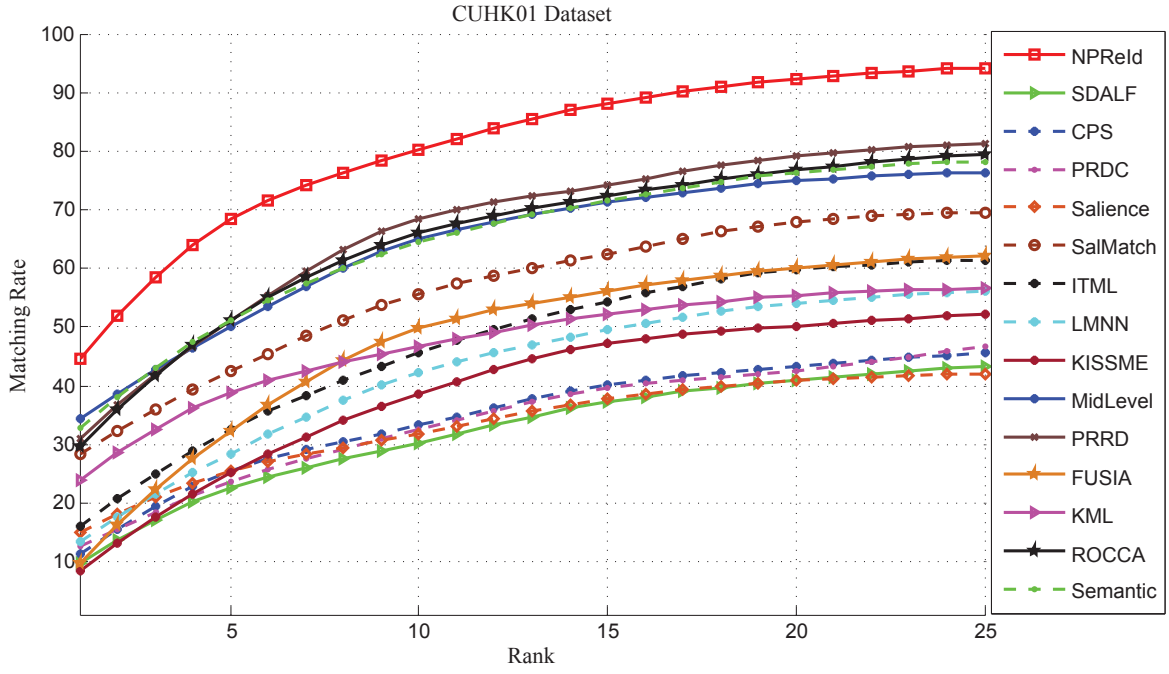

Figure 12: CMC curves for CUHK01 dataset

Table 4: Recognition rates (%) on the CUHK01 dataset with 485 image-pairs

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$
NPReId	44.52	68.31	80.31	92.35
Semantic [17]	32.70	51.20	64.40	76.30
ROCCA [18]	29.77	51.22	66.02	76.70
PRRD [19]	31.10	51.00	68.55	79.18
KML [12]	24.00	38.90	46.70	55.40
FUSIA [30]	9.80	32.40	49.80	60.10
MidLevel [13]	34.30	50.00	64.96	74.94
SalMatch [15]	28.45	42.50	55.68	67.95
Salience [14]	15.10	25.40	31.80	40.90
KISSME [10]	8.40	25.10	38.70	50.20
PRDC [9]	12.53	23.40	32.50	42.55
CPS [3]	11.35	25.56	33.23	43.35
SDALF [2]	9.90	22.60	30.30	41.00
LMNN [31]	13.45	28.50	42.25	54.11
ITML [32]	15.98	32.50	45.60	59.81

(a) (b)

Figure 13: (a) Image-pairs with drastic pose and viewpoint variations, (b) Image-pairs with improper background removal

learned model together with Z-score labelling is utilized to assign a reduced subspace of inlier clusters for a given probe. The principle of information gain is then suitably formulated to quantify each feature channel. The informative attributes are then incorporated in a correlation based distance measure to re-identify a probe within the look-alike inlier clusters. The tabular results alongside the performance curves on three benchmark datasets demonstrate the superiority of the proposed method over its counterparts.

Acknowledgment

This work is supported by Grant Number SB/FTP/ETA-0059/2014 by Science and Engineering Research Board (SERB), Department of Science & Technology, Government of India.

References

[1] Bryan Prosser, Shaogang Gong, and Tao Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *BMVC*, volume 8, pages 164–1. Citeseer, 2008.

[2] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2360–2367. IEEE, 2010.

[3] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom pictorial structures for re-identification. In *BMVC*, volume 1, page 6. Citeseer, 2011.

[4] Igor Kviatkovsky, Amit Adam, and Ehud Rivlin. Color invariants for person reidentification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1622–1634, 2013.

[5] Bingpeng Ma, Yu Su, and Frédéric Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014.

[6] Shi-Chang Shi, Chun-Chao Guo, Jian-Huang Lai, Shi-Zhe Chen, and Xiao-Jun Hu. Person re-identification with multi-level adaptive correspondence models. *Neurocomputing*, 168:550–559, 2015.

[7] Annan Li, Luoqi Liu, Kang Wang, Si Liu, and Shuicheng Yan. Clothing attributes assisted person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):869–878, 2015.

[8] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan 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*, pages 2197–2206, 2015.

[9] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 649–656. IEEE, 2011.

[10] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE, 2012.

[11] Zhen Li, Shiyu Chang, Feng Liang, Thomas Huang, Liangliang Cao, and John Smith. Learning locally-adaptive decision functions for person verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3610–3617, 2013.

[12] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision*, pages 1–16. Springer, 2014.

[13] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Learning mid-level filters for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 144–151. IEEE, 2014.

[14] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3586–3593. IEEE, 2013.

[15] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Person re-identification by salience matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2528–2535, 2013.

[16] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *Computer Vision–ECCV 2014*, pages 688–703. Springer, 2014.

[17] Liang Zheng, Shengjin Wang, Lu Tian, Fei He, Ziqiong Liu, and Qi Tian. Query-adaptive late fusion for image search and person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1741–1750, 2015.

[18] Le An, Songfan Yang, and Bir Bhanu. Person re-identification by robust canonical correlation analysis. *IEEE Signal Processing Letters*, 22(8):1103–1107, 2015.

[19] Le An, Mehran Kafai, Songfan Yang, and Bir Bhanu. Person reidentification with reference descriptor. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4):776–787, 2016.

[20] Stephen Grossberg. Neural facades: Visual representations of static and moving form-and-color-and-depth. *Mind & Language*, 5(4):411–456, 1990.

[21] Nebojsa Jojic, Alessandro Perina, Matteo Cristani, Vittorio Murino, and Brendan Frey. Stel component analysis: Modeling spatial correlations in image class structure. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2044–2051. IEEE, 2009.

[22] Joost Van De Weijer and Cordelia Schmid. Coloring local feature extraction. In *Computer Vision–ECCV 2006*, pages 334–348. Springer, 2006.

[23] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Computer Vision–ECCV 2008*, pages 262–275. Springer, 2008.

[24] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 3(5), 2007.

[25] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Asian Conference on Computer Vision*, pages 31–44. Springer, 2012.

[26] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research*, 3:583–617, 2003.

[27] Ofir Pele and Michael Werman. The quadratic-chi histogram distance family. In *European conference on computer vision*, pages 749–762. Springer, 2010.

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

[29] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary. Person re-identification by support vector ranking. In *BMVC*, volume 2, page 6, 2010.

[30] Ryan Layne, Timothy M Hospedales, and Shaogang Gong. Re-id: Hunting attributes in the wild. In *British Machine Vision Conference, Nottingham, England, 2014*. British Machine Vision Association, BMVA, 2014.

[31] Kilian Q Weinberger and Lawrence K Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(Feb):207–244, 2009.

[32] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007.

BibTeX

Cite this article as:
Aparajita Nanda, Pankaj K. Sa, Suman K. Choudhury, Sambit Bakshi, and Banshidhar Majhi.
A Neuromorphic Person Re-Identification Framework for Video Surveillance.
IEEE Access, 5:6471-6482, 2017. DOI: 10.1109/ACCESS.2017.2686438

```
@Article{NPReId,
author = {Aparajita Nanda and Pankaj K. Sa and Suman K. Choudhury and Sambit Bakshi and Banshidhar Majhi},
title = {A {N}euromorphic {P}erson {R}e-{I}dentification {F}ramework for {V}ideo {S}urveillance},
journal = {IEEE Access},
year = {2017},
volume = {5},
pages = {6471-6482},
doi = {10.1109/ACCESS.2017.2686438}
}
```