

Received February 14, 2017, accepted March 13, 2017, date of publication March 23, 2017, date of current version May 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2686438

A Neuromorphic Person Re-Identification Framework for Video Surveillance

**APARAJITA NANDA, PANKAJ KUMAR SA, SUMAN KUMAR CHOUDHURY,
SAMBIT BAKSHI, AND BANSHIDHAR MAJHI**

Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela 769008, India

Corresponding author: S. Bakshi (sambitbaksi@gmail.com)

This work was supported by the Science and Engineering Research Board, Department of Science and Technology, Government of India, under Grant SB/FTP/ETA-0059/2014.

ABSTRACT This paper presents a neuromorphic person re-identification (NPReId) framework to establish the correspondence among individuals observed across two disjoint camera views. The proposed framework comprises three modules (*observation*, *cognition*, and *contemplation*), inspired by the form-and-color-and-depth (FACADE) theory model of object recognition system. In the *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 the *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 toward defining a consensus cluster. Finally, in the *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 data sets: VIPER, iLIDS-VID, and CUHK01. The experimental results, with respect to recognition rates, demonstrate that the proposed framework can obtain superior performance as compared with the counterparts. The proposed framework, along with its low-rank bound property, further establishes its suitability in practical scenarios through yielding high cluster hit rate with low database penetration.

INDEX TERMS Surveillance, person re-identification, recognition, consensus clustering.

I. 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 Fig. 1. Use of biometric traits such as face, gait, periocular, fingerprint, etc. seems to be good candidates towards solving the classic person re-identification problem in context of visual surveillance. However, surveillance systems cannot bear the luxury of constrained environment where the images could be recorded as desired. The images are usually of very poor resolution hindering the application of biometric system in such scenario, especially in *online* operating mode of visual surveillance. In addition, person re-identification suffers from few severe challenges, such as, indistinguishable attire, unlike appearance, non-affine pose variations, varying background, partial occlusion etc; images in Fig. 2 show some of the typical scenarios where the above challenges can be well observed.

To address the above challenging issues, most of the existing literature focus on the following two areas — (i) a



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



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.

robust pedestrian signature in terms of invariant feature representation, and (ii) an efficient similarity measure to find the potential match. 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]. Liu *et al.* propose an unsupervised approach to learn the bottom-up feature importance [4]. Kviatkovsky *et al.* propose an invariant color signature in the log-chromaticity space by considering the color distribution under different lighting conditions [5]. Ma *et al.* suggest a biological covariance (BiCov) descriptor to address the problem with illumination change [6]. Shi *et al.* formulate a multi-level adaptive correspondence method for handling the misalignment of body parts [7]. In another work, a part based segmentation approach is suggested to solve the problem with pose misalignments [8]. 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 [9]. Zheng *et al.* [10] introduce a probabilistic relative distance comparison 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 [11]. Li *et al.* model a quadratic

decision function for metric learning [12]. In another work, a kernel-based distance learning approach is presented to improve the re-identification accuracy [13]. Subsequently, the discriminative and representative patches are collected for feature learning [14]. Zaho *et al.* propose an unsupervised salience learning model to learn the salience regions of human appearance [15], [16]. Wang *et al.* design a video ranking model by simultaneously selecting and matching the reliable space-time features from the image sequence [17]. Zheng *et al.* propose a score-level fusion scheme that automatically selects an appropriate set of features from the unlabeled data [18]. An *et al.* formulate a robust canonical correlation analysis to map the samples from two disjoint views into a subspace followed by similarity matching [19], [20]. Shen *et al.* introduce a boosting-based approach for learning the correspondence patch-matching probabilities between the image pairs [21].

It has been observed from the above-mentioned approaches that the pedestrian signature with a 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 an exhaustive search of the 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. We aim to partially and maximally satisfy two contradictory criteria: (i) to achieve high recognition rate, and (ii) to maintain low penetration of dataset for fast processing. To simultaneously achieve these two aspects, we design a person re-identification framework by following the principles of FACADE (Form-And-Color-And-Depth) theory [22].

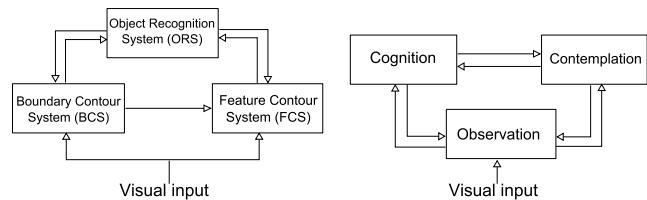


FIGURE 3. The FACADE modules. Left: the BCS-FCS includes the background segmentation, foreground feature extraction, which are further recognized by the ORS. Right: the interactive modules of proposed re-identification system (NPReId) based on FACADE theory.

FACADE, 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 Fig. 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

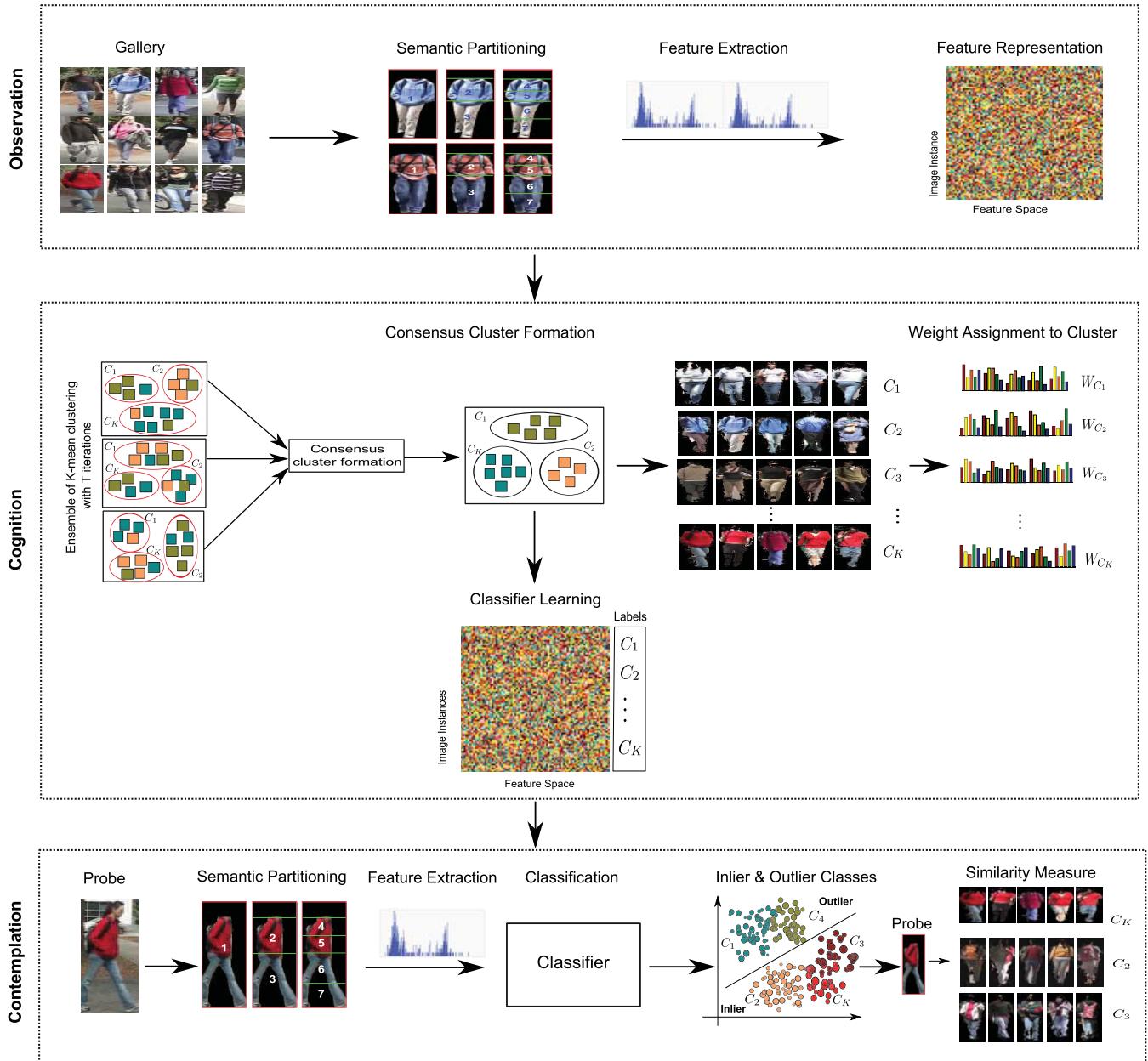


FIGURE 4. Overview of the proposed NPrReId framework with three interactive modules – *Observation*, *Cognition*, and *Contemplation* along with the steps in each module.

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. Based on the above idea, we design a neuromorphic person re-identification system (NPrReId) following the FACADE theory that comprises three interactive modules – *observation*, *cognition*, and *contemplation*. 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.

The rest of this article is organized as follows. The proposed NPrReId system is elaborated in Section II. Experimental results on standard datasets along with other state-of-the-art methods are presented in Section III. Finally, concluding remarks are provided in Section IV.

II. PROPOSED NPrReId FRAMEWORK

In this paper, we present the NPrReId framework, in line with the FACADE theory, to establish the correspondence between a probe and a subset of gallery images. Fig. 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 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. Proposed NPReId framework, through mentioned three modules, strives for—(i) selection of complementary features combination, (ii) formation of inlier subset of gallery candidates for a given probe, where the probability of finding the match is very high.

A. OBSERVATION

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

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 [23], a preprocessing operation, to suppress the background content prior to semantic partitioning.

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 [24] 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 Fig. 5.

2) FEATURE EXTRACTION

In person re-identification, complementary appearance cues need to be integrated to generate a robust feature

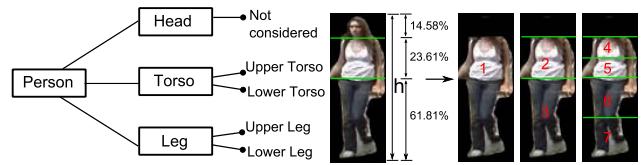


FIGURE 5. Semantic partitioning of the body structure.

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-Saturation¹ (HwS) [25] and CbCr, the intensity channel includes Y, and the texture channel, adapted from [26], 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 [27], iLIDS-VID [17], and CUHK01 [28] are taken into consideration for this experimental purpose. The details of these datasets are given in Section III-A.

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 Fig. 6(a). 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 Fig. 6(b). The other two datasets also result in similar observations as shown in Fig. 6(c) and 6(d).

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 $d = 2688$.

B. COGNITION

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

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

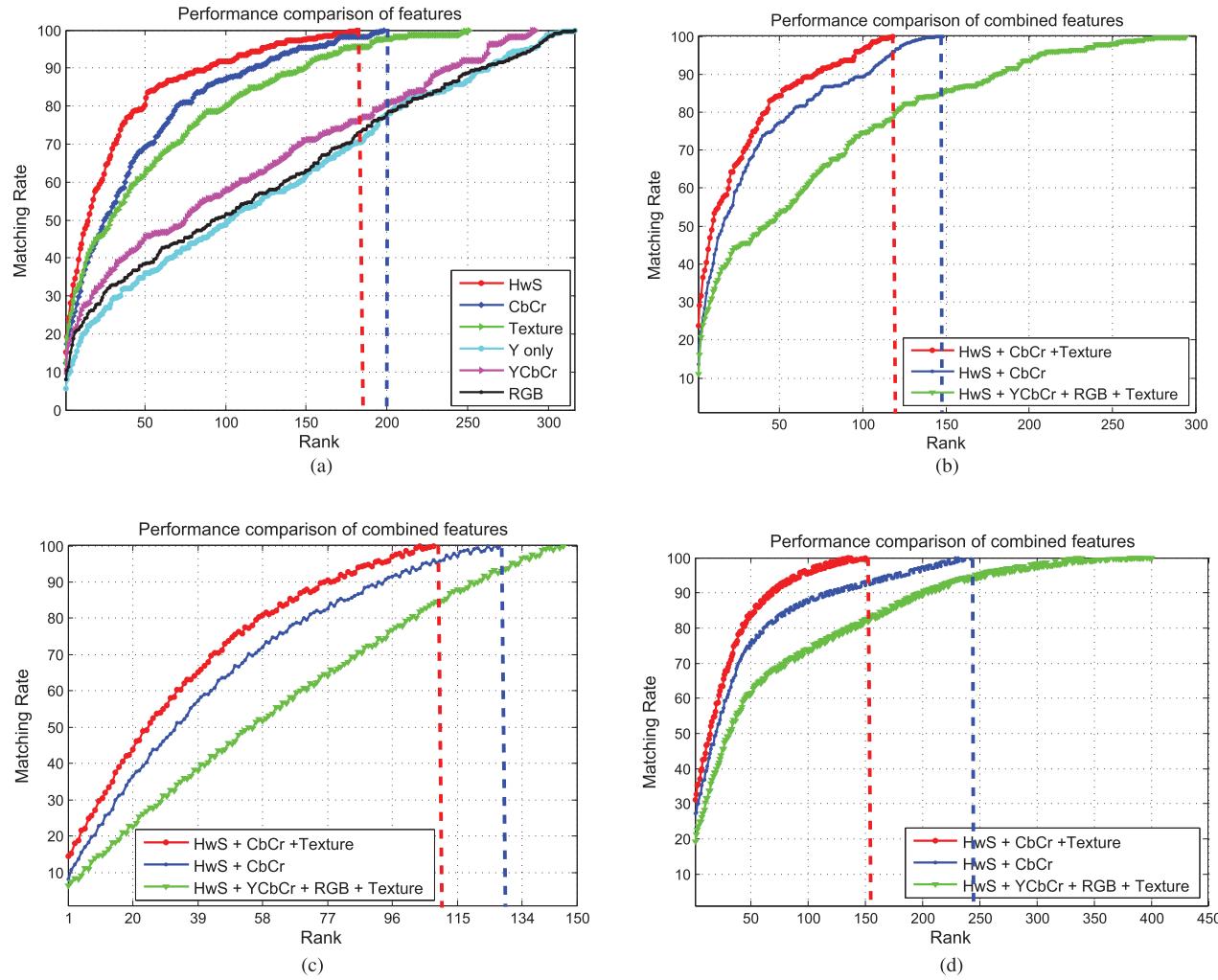


FIGURE 6. Comparative analysis through CMC curves. (a) CMC curve across each individual feature channel for VIPeR. (b) CMC curve across various features combination for VIPeR. (c) CMC curve across various features combination for iLIDS-VID. (d) CMC curve across various features combination for CUHK01.

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. The value of K is estimated by using the self-tuning algorithm [29]. 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.

- We then apply the consensus based meta-graph clustering algorithm (MCLA) [30], an approach to re-cluster the clusters, to merge the $T \times K$ clusters into K consensus clusters $\{C_1, C_2, \dots, C_K\}$.
- A classifier model is then built using support vector machine (SVM) of Gaussian kernel to learn the relationship between the gallery feature vectors $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$ and their corresponding cluster labels $C = \{C_1, C_2, \dots, C_K\}$. This learned model has latter been used in the contemplation stage of the framework.

2) WEIGHT ASSIGNMENT TO EACH OF THE CONSENSUS CLUSTER

This section analyses each feature channel relative to their contribution towards defining a cluster. We apply information gain principle to quantify each feature channel.

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

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

Each feature vector I_i is represented with f feature channels ($I_i = \{I_i^\alpha\}$, $\alpha = 1, 2, \dots, f$). According to the principle of information gain, the contribution of a feature channel α with respect to a training pair ϕ can be expressed through Eq. 2.

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

where $H(\phi)$ measures the entropy of the training set; $E(\alpha)$ denotes the set of all possible values of a feature channel α . The above expression, for all possible values of α yields a f -dimensional vector $L = \{l_1, l_2, \dots, l_f\}$ that signifies the relative contribution of each feature channel towards defining a consensus cluster. It can be observed that larger-contributory feature channels highlight the commonality among the images within a consensus cluster. In other words, the low contributory feature channels are more informative in distinguishing the images within the same cluster. Accordingly, while searching a probe within the look-alike of a consensus cluster, assignment of higher priority to feature channels with low contribution becomes an obvious choice. Hence the above vector L is complemented to represent the required weight vector $W = \{w_1, w_2, \dots, w_f\}$ as shown in Eq. 3.

$$w_t = \sum_{j=1}^f l_j - l_t / \sum_{i=1}^f \left(\sum_{j=1}^f l_j - l_i \right) \quad (3)$$

The weight vector $W = \{w_1, w_2, \dots, w_f\}$ is essential for the similarity measure and employed in the Contemplation module.

C. CONTEMPLATION

In this module, the learned classification model, as discussed in Section II-B1, is employed to find a set of inlier consensus clusters for a given probe. A correlation based weighted similarity measure is then applied to find the exact match within the set of inlier clusters. The subsequent paragraphs detail both the steps.

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

Algorithm 1 Inlier Cluster Identification

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

```

1  $\tilde{\mathcal{C}} \leftarrow \phi$ ; // Initialize  $\tilde{\mathcal{C}}$  with empty set
2 Apply probe feature vector  $I_p$  on classification model M that results in  $K$  classification scores
    $S = \{s_1, s_2, \dots, s_K\}$  with respect to the set of consensus clusters  $\mathcal{C} = \{C_1, C_2, \dots, C_K\}$ ;
3  $s_{\max} \leftarrow \max(S)$ ; // Extract the maximum score from  $S$ 
4  $\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} \cup \{C_j\}$ , where  $s_j == s_{\max}$ ;  

   // include the cluster with maximum classification score in  $\tilde{\mathcal{C}}$ 
5  $S \leftarrow S - \{s_{\max}\}$ ; // Exclude  $s_{\max}$  from  $S$ 
6 Create a vector  $A$  to store  $m$  random numbers ( $m \geq 10$ ) following a normal distribution with mean  $\mu = s_{\max}$  and standard deviation  $\sigma = 0.1$ ;  

   //  $A$  is a vector with no outlier samples  

   //  $\sigma$  is empirically set to 0.1
7 for  $j \leftarrow 2$  to  $K$ , do
8    $s_{\max} \leftarrow \max(S)$ ; // Extract the next maximum score from  $S$ 
9    $Z_{\mu_A, \sigma_A} \leftarrow (s_{\max} - \mu_A) / \sigma_A$ ;  

   //  $\mu_A$  and  $\sigma_A$  are the mean and standard deviation of  $A$ 
10  if  $|Z_{\mu_A, \sigma_A}| \leq 3$  then
    // any value with Zscore>3 and Zscore<-3 is usually set as outlier, where 3 is the empirically set threshold
     $\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} \cup \{C_j\}$ , where  $s_j == s_{\max}$ ;  

     $A \leftarrow A \cup \{s_{\max}\}$ ; // include  $s_{\max}$  in  $A$ 
     $S \leftarrow S - \{s_{\max}\}$ ;
11 else
12   break;
```

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 identify the set of inlier clusters for a given probe, enumerated in Algorithm 1.

1) RE-IDENTIFICATION RANKING BY SIMILARITY MEASURE

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

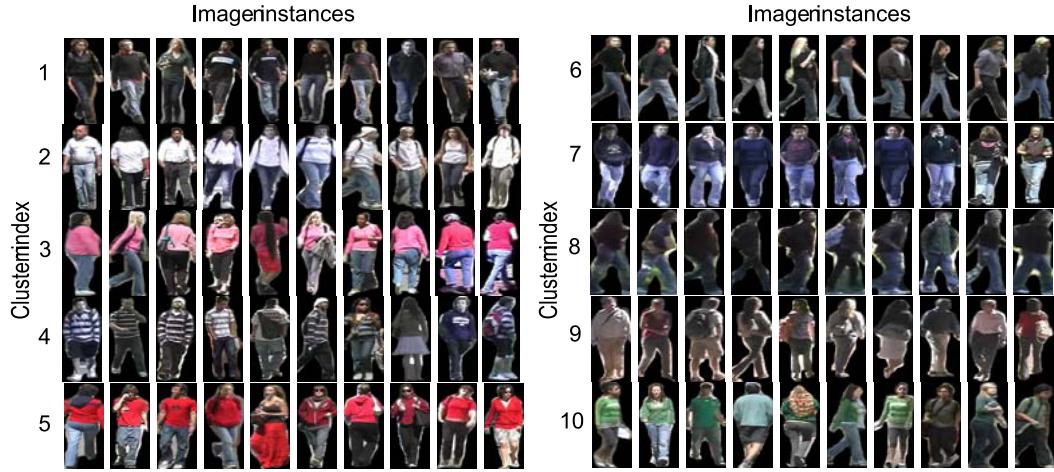


FIGURE 7. Sample consensus clusters in VIPeR dataset with unlabeled pedestrians.

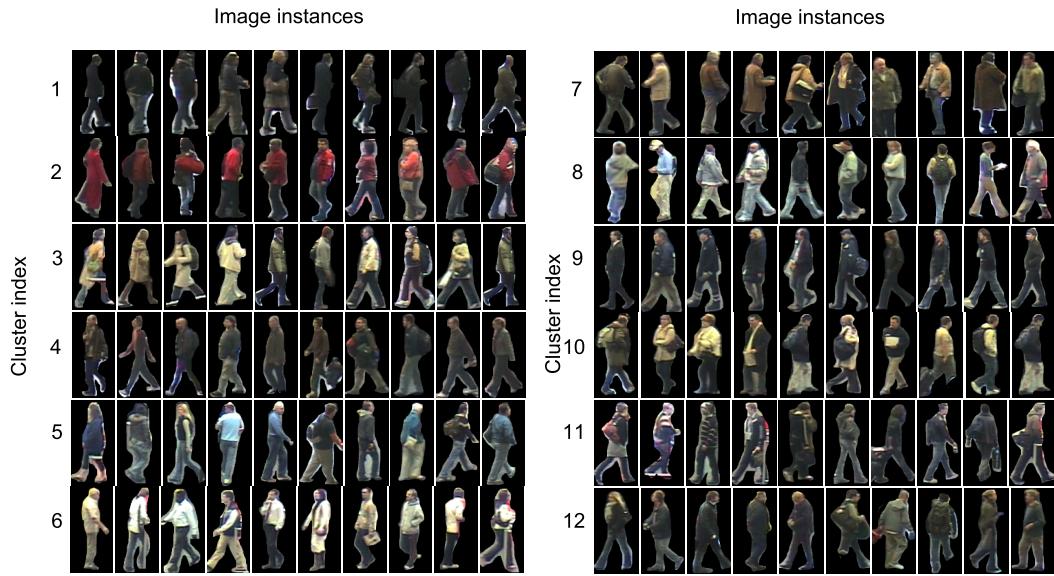


FIGURE 8. Sample consensus clusters in iLIDS-VID dataset with unlabeled pedestrians.

(χ_{quad}) [31] 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 $I_p = \{U_1, U_2, \dots, U_f\}$ and gallery feature vector $I_g = \{V_1, V_2, \dots, V_f\}$ is given by Eq. 4.

$$\mathcal{D}(I_p, I_g) = \sum_{i=1}^f w_i \cdot \chi_{quad}(U_i, V_i) \quad (4)$$

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

III. 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 compared with some of the state-of-the-art methods and the efficacy of inlier set is validated for all the datasets. 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.

A. DATASETS AND STATE-OF-THE-ART METHODS

We evaluate our propose framework on three benchmark datasets namely, *Viewpoint Invariant Pedestrian Recognition* (VIPeR, [27]), *iLIDS Video re-IDentification Dataset* (iLIDS-VID, [17]), and *Campus dataset* (CUHK01, [28]). These datasets are publicly available and widely used in

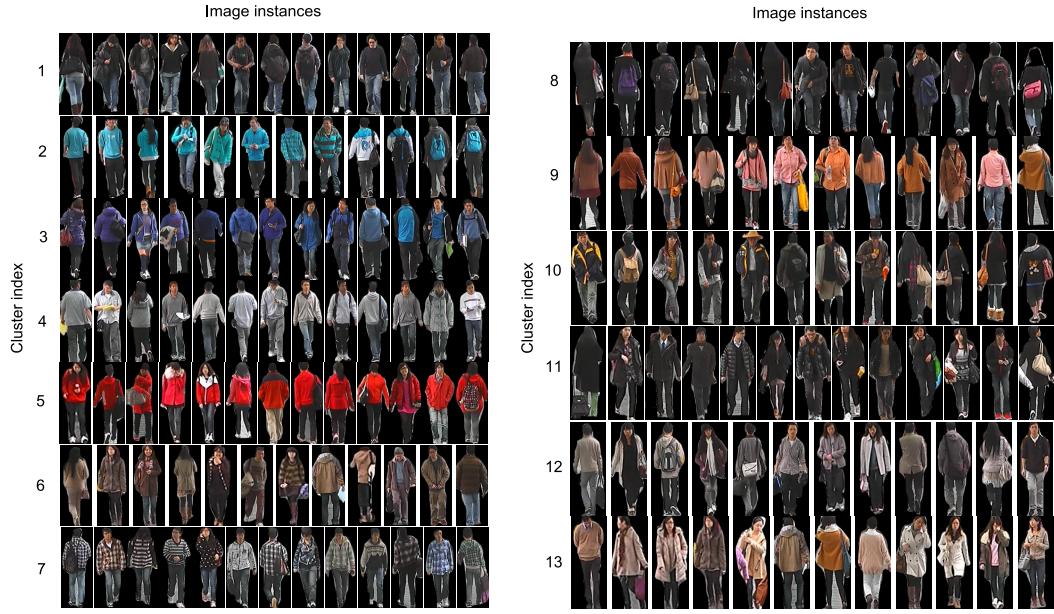


FIGURE 9. Sample consensus clusters in CUHK01 dataset with unlabeled pedestrians.

TABLE 1. Description of datasets used.

Name	# images	Resolution	Challenges
VIPeR [27]	1264 (632 image pairs)	128 × 48 24-bit RGB	Viewpoint variation Illumination change
iLIDS-VID [17]	600 (300 image pairs)	128 × 64 24-bit RGB	Illumination change Similarity in clothing cluttered background Partial occlusion
CUHK01 [28]	1942 (971 image pairs)	160 × 60 24-bit RGB	Pose variations Illumination change

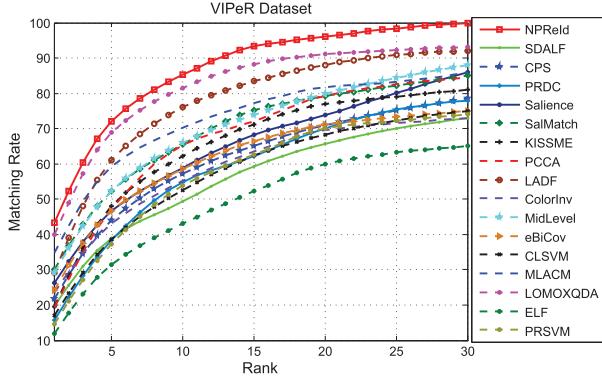


FIGURE 10. CMC curves for VIPeR dataset.

the person re-identification task. The image pairs of the datasets are captured by multiple camera views at different locations in different instances of time. 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 LOMOXQDA [9], MLACM [7], eBiCov [6], CLSVM [8],

MidLevel [14], LADF [12], ColorInv [5], SalMatch [16], Salience [15], KISSME [11], PCCA [32], PRDC [10], CPS [3], SDALF [2], ELF [26], and PRSVM [33]. We have chosen the following methods for the iLIDS-VID dataset: the supervised approach (MidLevel [14], PRDC [10]), unsupervised approach (Salience [15], SalMatch [16], CPS [3]) and multi-shot approaches (MSSDALF [2], MSColor RSVM [17], MSColorLBPRSVM [17]). Similarly, the simulated methods chosen for CUHK01 dataset are: Semantic [18], ROCCA [19], PRRD [20], KML [13], FUSIA [34], MidLevel [14], SalMatch [16], Salience [15], KISSME [11], PRDC [10], CPS [3], SDALF [2], LMNN [35], and ITML [36].

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 Fig. 7, 8, and 9; the appearance similarity among the members of each consensus cluster is very much apparent in these figures.

B. RESULTS ANALYSIS

Our experiments on the three datasets follow the evaluation protocol in [26]. 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 trial, One of each image

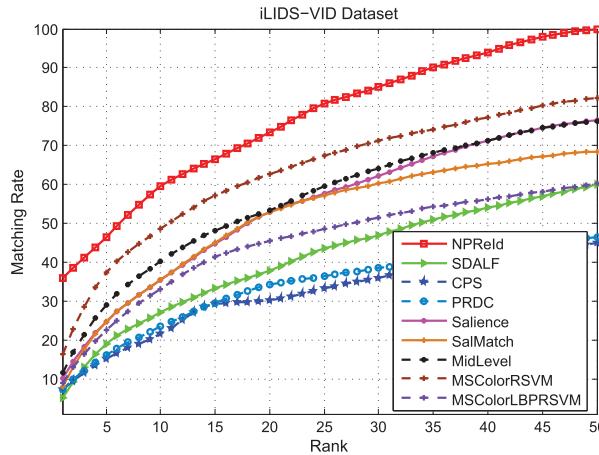


FIGURE 11. CMC curves for iLIDS-VID dataset.

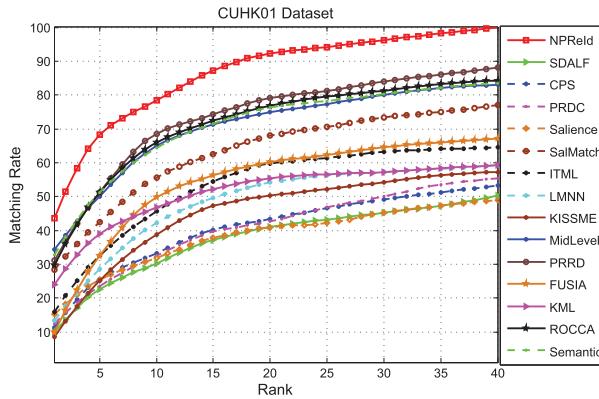


FIGURE 12. CMC curves for CUHK01 dataset.

TABLE 2. Recognition rates (%) on the VIPeR dataset with 316 image-pairs.

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 30$
LOMOXQDA [9]	40.00	68.90	81.50	91.10	93.25
MLACM [7]	34.87	59.27	70.19	81.77	86.21
eBiCov [6]	24.34	46.75	58.48	71.17	75.21
CL SVM [8]	17.09	38.61	52.53	68.35	74.82
MidLevel [14]	29.10	52.30	65.90	79.90	88.11
LADF [12]	29.30	61.00	76.00	88.10	92.31
ColorInv [5]	23.51	43.04	55.16	69.59	73.25
Salience [15]	26.31	46.60	58.86	72.77	86.10
SalMatch [16]	30.16	52.31	65.54	79.15	84.91
PCCA [32]	19.27	47.00	65.00	79.00	84.50
KISSME [11]	19.60	47.00	62.60	78.00	81.10
PRDC [10]	15.66	38.42	53.86	70.09	78.12
CPS [3]	21.84	44.00	57.21	71.00	78.23
SDALF [2]	19.87	38.90	49.37	65.73	73.11
NPReId	43.36	72.12	85.21	96.05	100.00

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 characteristic (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 Fig. 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 3. Recognition rates (%) on the iLIDS – VID dataset with 150 image-pairs.

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 50$
MScolorRSVM [17]	16.40	37.30	48.50	62.60	82.31
MScolorLBRSVM [17]	20.00	44.00	52.70	68.00	91.20
MSSDALF [2]	5.10	19.00	27.10	37.90	60.21
MidLevel [14]	11.70	29.00	40.30	53.40	76.22
SalMatch [16]	8.01	24.80	35.40	52.90	68.51
Salience [15]	10.12	24.82	35.45	52.92	76.51
PRDC [10]	7.44	16.21	23.44	34.17	46.35
CPS [3]	7.32	15.31	21.52	30.20	45.28
NPReId	35.87	46.52	59.50	72.38	100.00

TABLE 4. Recognition rates (%) on the CUHK01 dataset with 486 image-pairs.

Method	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 40$
Semantic [18]	32.70	51.20	64.40	76.30	83.50
ROCCA [19]	29.77	51.22	66.02	76.70	84.23
PRRD [20]	31.10	51.00	68.55	79.18	88.21
KML [13]	24.00	38.90	46.70	55.40	59.25
FUSIA [34]	9.80	32.40	49.80	60.10	67.22
MidLevel [14]	34.30	50.00	64.96	74.94	83.24
SalMatch [16]	28.45	42.50	55.68	67.95	77.01
Salience [15]	15.10	25.40	31.80	40.90	49.13
KISSME [11]	8.40	25.10	38.70	50.20	57.21
PRDC [10]	12.53	23.40	32.50	42.55	55.26
CPS [3]	11.35	25.56	33.23	43.35	53.23
SDALF [2]	9.90	22.60	30.30	41.00	50.21
LMNN [35]	13.45	28.50	42.25	54.11	59.35
ITML [36]	15.98	32.50	45.60	59.81	64.50
NPReId	44.52	68.31	80.31	92.35	100.00

It can be observed that the proposed NPReId framework, at rank-01 possesses at least 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.

C. INLIER SET VALIDATION

We validate the efficacy of the inlier set using two parameters *Penetration Rate (PR)* and *Cluster Hit (CH)*. The *PR* is defined as the expected ratio of clusters searched to the total number clusters within the gallery set for a successful match. The *CH* is defined as the percentage of probes that are successfully searched at the corresponding clusters. Mathematically, *PR* and *CH* are defined as

$$PR = \frac{m}{K} \quad (5)$$

where m : number of clusters have been searched for finding the true image pair, K : total number of clusters.

$$CH = \frac{M_{I_p}}{N_{I_p}} \quad (6)$$

M_{I_p} : number of probe images have been successfully found in the corresponding clusters, N_{I_p} : total number of images in the probe set. Fig. 13(a), 13(b), and 13(c) illustrate the inlier set validation on three datasets. The lower penetration rate signifies the reduction in the number of clusters to be searched with respect to the total number of clusters. The results demonstrate that in VIPeR 89%, in iLIDS-VID 83%, and in CUHK01 92% of the probe are successfully searched in the inlier set.

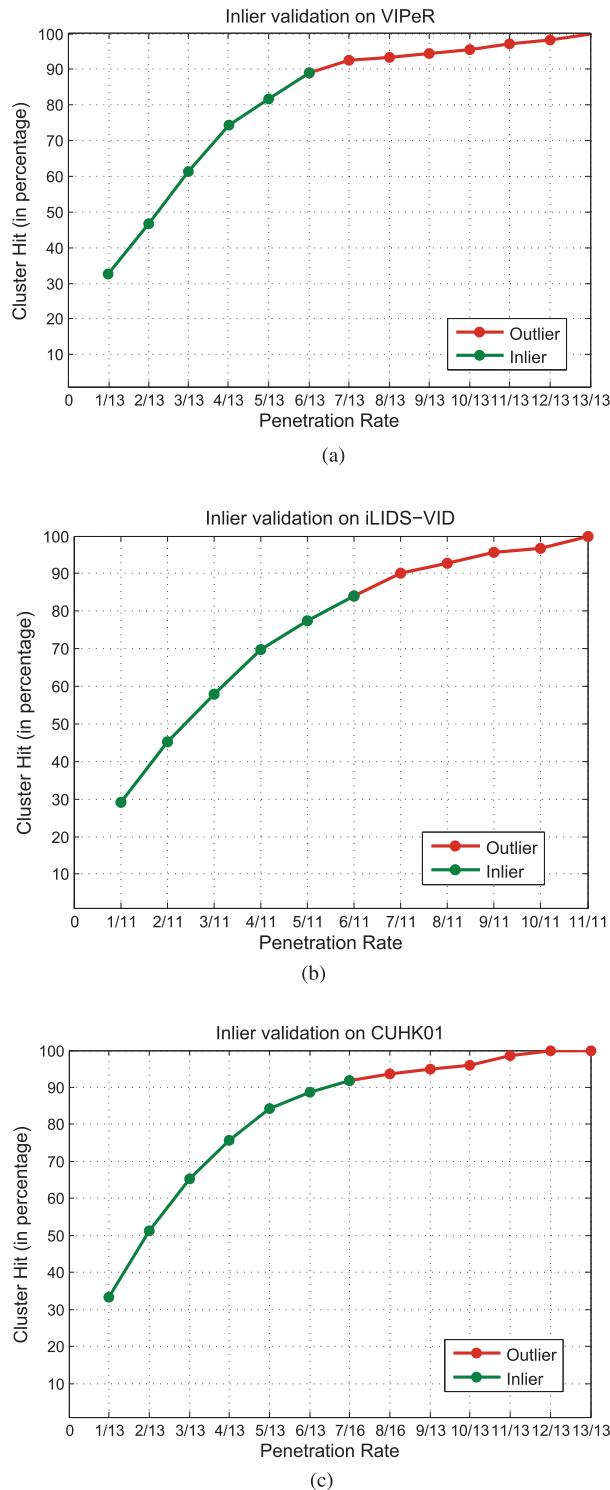


FIGURE 13. Inlier set validation on different datasets. (a) VIPeR dataset with $K = 13$. (b) iLIDS-VID dataset with $K = 11$. (c) CUHK01 dataset with $K = 13$.

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 Fig. 14(a).



FIGURE 14. Some challenging scenarios. (a) Image-pairs with drastic pose and viewpoint variations. (b) Image-pairs with improper background removal.

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.

- (ii) Cluttered background in the bounding box of a pedestrian may lead to overfitting. 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 Fig. 14(b).

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. Further, person re-identification possess many more interesting challenges under the mobile environment, particularly with the advancement of newer technologies like mobile video surveillance [37] and wireless video surveillance [38].

IV. 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 learned model together with Z-score labeling 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 channels are then incorporated in a correlation based distance measure to re-identify a probe within the look-alike inlier clusters. The results recorded, alongside the performance curves on three benchmark datasets, demonstrate the superiority of the proposed method over state-of-the-art methods. Analysing theoretically, the gain in performance owes to (i) semantic partitioning of body structure, (ii) consensus cluster formation, and (iii) intelligent discovery of inlier set.

ABBREVIATIONS

BCS	: Boundary Contour System
CH	: Cluster Hit
CLT	: Central Limit Theorem
CMC	: Cumulative Matching Characteristics
FACADE	: Form-and-Color-and-Depth
FCS	: Feature Contour System
FoV	: Field of View
HwS	: Hue-weighted-Saturation
MCLA	: Meta-graph CLustering Algorithm
NPReId	: Neuromorphic Person Re-Identification
ORS	: Object Recognition System
PR	: Penetration Rate
VIPeR	: Viewpoint Invariant Pedestrian Recognition

REFERENCES

- [1] B. Prosser, S. Gong, and T. Xiang, "Multi-camera matching using bi-directional cumulative brightness transfer functions," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, vol. 8. 2008, pp. 1–164.
- [2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 2360–2367.
- [3] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2011, vol. 1, no. 2, p. 6.
- [4] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important?" in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 391–401.
- [5] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1622–1634, Jul. 2013.
- [6] B. Ma, Y. Su, and F. Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification," *Image Vis. Comput.*, vol. 32, nos. 6–7, pp. 379–390, 2014.
- [7] S.-C. Shi, C.-C. Guo, J.-H. Lai, S.-Z. Chen, and X.-J. Hu, "Person re-identification with multi-level adaptive correspondence models," *Neurocomputing*, vol. 168, pp. 550–559, Nov. 2015.
- [8] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing attributes assisted person reidentification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 869–878, May 2015.
- [9] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 2197–2206.
- [10] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 649–656.
- [11] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2288–2295.
- [12] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 3610–3617.
- [13] F. Xiong, M. Gou, O. Camps, and M. Sznajer, "Person re-identification using kernel-based metric learning methods," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 1–16.
- [14] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 144–151.
- [15] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 3586–3593.
- [16] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2528–2535.
- [17] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 688–703.
- [18] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1741–1750.
- [19] L. An, S. Yang, and B. Bhanu, "Person re-identification by robust canonical correlation analysis," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1103–1107, Aug. 2015.
- [20] L. An, M. Kafai, S. Yang, and B. Bhanu, "Person reidentification with reference descriptor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 776–787, Apr. 2016.
- [21] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, "Person re-identification with correspondence structure learning," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 3200–3208.
- [22] S. Grossberg, "Neural facades: Visual representations of static and moving form-and-color-and-depth," *Mind Lang.*, vol. 5, no. 4, pp. 411–456, 1990.
- [23] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey, "Stel component analysis: Modeling spatial correlations in image class structure," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 2044–2051.
- [24] L. da Vinci, *Da Vinci Notebooks*. London, U.K.: Profile, 2006, pp. 1–224.
- [25] J. van de Weijer and C. Schmid, "Coloring local feature extraction," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2006, pp. 334–348.
- [26] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 262–275.
- [27] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. Int. Workshop Perform. Eval. Tracking Survell. (PETS)*, 2007, vol. 3, no. 5, pp. 1–7.
- [28] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2012, pp. 31–44.
- [29] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1601–1608.
- [30] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
- [31] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 749–762.

- [32] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 2666–2672.
- [33] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2010, vol. 2, no. 5, p. 6.
- [34] R. Layne, T. M. Hospedales, and S. Gong, "Re-ID: Hunting attributes in the wild," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2014, p. 6.
- [35] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.
- [36] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 209–216.
- [37] M. Abu-Lebdeh, F. Belqasmi, and R. Glitho, "An architecture for QoS-enabled mobile video surveillance applications in a 4G EPC and M2M environment," *IEEE Access*, vol. 4, pp. 4082–4093, 2016.
- [38] Y. Ye, S. Ci, A. K. Katsaggelos, Y. Liu, and Y. Qian, "Wireless video surveillance: A survey," *IEEE Access*, vol. 1, pp. 646–660, Sep. 2013.



APARAJITA NANDA received the master's degree from the Sambalpur University Institute of Information Technology, in 2011. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, National Institute of Technology at Rourkela, Rourkela, India. Her research interests include computer vision, visual surveillance, machine learning, and data mining.



PANKAJ KUMAR SA received the Ph.D. degree in computer science from NIT Rourkela, in 2010. He is currently an Assistant Professor with the CSE Department, NIT Rourkela. He has authored or co-authored a number of research articles in various journals, conferences, and book chapters. His research interests include computer vision, biometrics, and visual surveillance. He is a member of the CSI. He has co-investigated some R&D projects funded by SERB, PXE, DeitY, and ISRO. He has been conferred with various prestigious awards and honors. Apart from research and teaching, he is also actively involved with the automation of NIT Rourkela, where he conceptualizes and engineers the automation process.



SUMAN KUMAR CHOUDHURY received the M.Tech. degree from the National Institute of Technology at Rourkela, Rourkela, India, in 2013, where he is currently pursuing the Ph.D. degree. His research interests include computer vision, video surveillance, image processing, and pattern recognition.



SAMBIT BAKSHI received the master's degree, and the Ph.D. degree in computer science and engineering, in 2015. He is currently with the Centre for Computer Vision and Pattern Recognition, National Institute of Technology at Rourkela, Rourkela, India, where he also serves as an Assistant Professor with the Department of Computer Science and Engineering. He is a member of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence. He received the prestigious Innovative Student Projects Award 2011 from the Indian National Academy of Engineering for his master's thesis. He has been serving as an Associate Editor of the *International Journal of Biometrics*, since 2013. He serves as an Associate Editor of the IEEE ACCESS, the *PLOS One*, the *Innovations in Systems and Software Engineering* (NASA Journal), and the *International Journal of Biometrics*. He has over 30 publications in journals, reports, and conferences.



BANS HIDHAR MAJHI is currently a Professor of the Department of Computer Science and Engineering, National Institute of Technology at Rourkela, Rourkela, India, where he is also serving as the Dean (Academic). He is the Chairman and the Principal Investigator with the Centre for Computer Vision and Pattern Recognition, National Institute of Technology at Rourkela, Rourkela. He serves as an Editor and Reviewer for many journals and conferences. He has over 100 publications in reputed conferences, journals, and book chapters.

• • •