IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 11, 2019, accepted December 2, 2019, date of publication December 10, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958911

Target Field of View Prediction Using Artificial Pheromones for People Reidentification

EVERARDO SANTIAGO-RAMIREZ[1], JOSE ANGEL GONZALEZ-FRAGA[2], EVERARDO GUTIERREZ LOPEZ[2], OMAR ALVAREZ-XOCHIHUA[2], AND JUAN ACOSTA-GUADARRAMA[1]

[1]Instituto de Ingeniería y Tecnología, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez 32310, México
[2]Facultad de Ciencias, Universidad Autónoma de Baja California, Ensenada 22860, México

Corresponding author: Everardo Santiago-Ramirez (everardo.santiago@uacj.mx)

The work of E. Santiago-Ramirez was supported in part by the Programa para el Desarrollo Profesional Docente (PRODEP) under Project UACJ-PTC-424.

ABSTRACT People reidentification is a fundamental task in automated video surveillance based on computer vision. Reidentification happens when a person seen in a field of view is the same that has been observed in other fields of view. A person who has disappeared from one field of view can appear in any other within a camera network. Instead of looking for the person in all neighboring fields of view, for an intelligent video surveillance system, it is more practical to predict which of the neighboring camera views the person could appear. This prediction can become achieved by learning the paths the person usually follows in the camera network. The ant colony optimization technique has properties that can get exploited for this purpose; precisely, the accumulation and evaporation of artificial pheromones are used to learn the paths. After the learning process, the proposed method can make predictions every time that the person leaves a field of view. Such prediction is evaluated to obtain feedback and further tune the learning process. The path followed by the person becomes obtained by tracking their face image within and between fields of view using correlation filters as descriptors. The results obtained from an extensive experiment show that the field of view that the person selects to visit can be successfully predicted using artificial pheromones, and thus, reduce the resources that require reidentification.

INDEX TERMS People reidentification, ant colony optimization, correlation filters.

I. INTRODUCTION

In most cities, there are cameras interconnected for video surveillance purposes, generating massive amounts of visual and non-visual data from that network. That makes it impossible for human operators to analyze and use the data efficiently. They could use it to, for example, learn the path that a person follows in the network and then use that data to predict in which view field the person could appear. The development of new artificial vision algorithms and computers that allow parallel processing of images makes it possible to analyze such data to obtain meaningful information automatically.

The task of following and authenticating a person in a network of disjoint cameras is known as reidentification (ReID). It determines whether the person viewed in a field of view (FoV) of a camera is the same person observed in other FoVs. ReID has a wide range of applications, such as public safety, long-term multicamera tracking, and forensic search [1]–[6].

ReID is an artificial-vision technique that includes detection, tracking, matching, and, for the biometric context, recognition. A person must be tracked within a field of view and between fields of view. Within a field of view, the tracking can get performed with acceptable results, by using existing approaches. Tracking between FoVs, on the other hand, remains an open problem because the appearance of the person can drastically change from a camera to another under real-world conditions. Thus, ReID requires the design of descriptors that can deal with appearance changes, variable illumination, occlusion, background clutter, poses, scale, and some other variations [4], [7], [8].

Generally, a person ReID approach can get split into four modules:

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyu Zhou.

1) *Person detection*. Given an arbitrary image, the goal of person detection is to determine whether there is any person in the image. If so, return the location and dimension of the region of interest.
2) *Person tracking*. Estimate the location of the previously detected person in each video frame.
3) *Descriptor design*. Data obtained by detection and tracking modules are used to segment the person's image for generating the descriptor, which can get constructed from data/cues as face [9]–[13]; visual appearance of the whole body [8], [14]–[21]; walking pattern [22], [23]; height and build [1]; and head, torso and limbs of a person [24], [25]; or a combination of these cues.
4) *Descriptors matching*. Descriptors created in disjoint FoVs get matched to determine whether they belong to the same person.

Modules 1 and 2 are performed on an FoV to obtain a set of images to build a descriptor for the person (Module 3). Module 4 can operate in two ways: 1) Compare descriptors generated from stored videos that became previously captured from the camera network, and 2) Generate and compare descriptors while the person goes through the camera network. This second method allows the implementation of a semi-real time system. With that, it is straightforward to notice that ReID works for both short- and long-term periods.

The semi-real time ReID task is computationally expensive. Let Fig. 1 represent a camera network where the red arrows mark the path $R = \{i, j, k, f\}$ followed by a person. From that path, it is possible to obtain in each FoV a video where the person to re-identify appears. A person who leaves the FoV i can appear in either j, k, or l. Searching the person in the whole neighborhood requires many computational resources and may require the intervention of an operator.

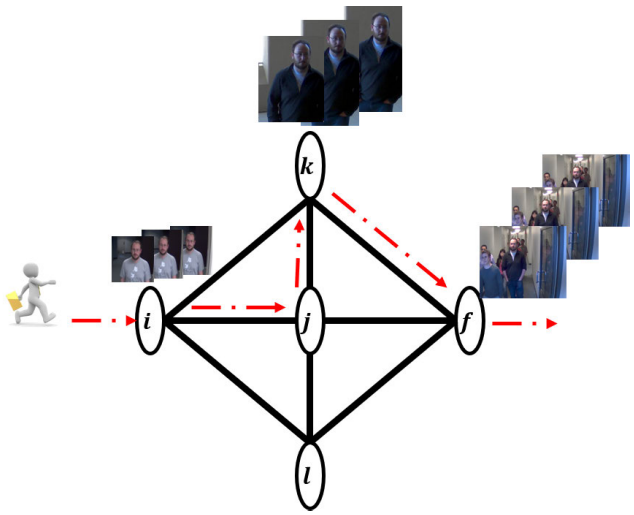

FIGURE 1. Camera network represented by a graph, with the fields of view f, i, j, k, and l; where i and j are the network's input and output, respectively.

Furthermore, each FoV is required to perform the person detection module, tracking module, and matching module on a sequence of images. This problem produces the need to predict the FoV that the person would select, which we call the target FoV in our research.

The target FoV prediction task is a new problem not ad-dressed within the reviewed literature. Addressing it in this work can contribute to attracting the attention of other researchers to this problem. This task has the potential to extend the capacity of ReID algorithms by allowing the development of systems that do not require the intervention of a human operator.

The target FoV prediction problem is addressed in this paper by an ant colony optimization (ACO) [26]-based method that learns the path, usually followed by the person in a camera network. The learned path is used to predict the fields of view that the person could visit during a semi-real time reidentification. Because surveillance systems under real-world conditions usually capture only part of the person, the descriptor for the proposed method is a composite correlation filter [20], [27] that combines the facial images of the same person into a single signal. Experimental results show that the proposed method can efficiently predict the target FoV, improve the speed of semi-real time reidentification, and reduce the use of computational resources.

The rest of this paper consists of the following. Section II briefly describes some existing approaches for person reidentification. Section III describes, in detail, the proposed method for predicting the FoVs that the person could visit. Section IV presents a discussion of the results obtained in real and simulated scenarios. Finally, Section V summarizes the main conclusions of this research.

II. RELATED WORK

No approaches were found that address the problem of predicting the target field of view in the reviewed literature. Therefore, this section describes approaches for camera network topology inference applied to ReID, and others based on learning models and traditional techniques.

A large number of cameras installed at universities, shopping centers, public squares, parks, among other places, have the purpose of, for example, tracking people, and incident prevention. Those cameras are interconnected in such a way that they form a network and maintain a spatio-temporal relationship known as camera network topology. Intelligent video surveillance based on computer vision requires the installation of a camera network in such a way as to efficiently extract useful information from a large number of videos [28], [29].

There are some proposed methods in the literature for camera-network topology inference applied to ReID. In [30], they proposed a framework that addresses both the person reidentification and camera network topology inference. First, a random forest-based classifier becomes trained to re-identify people. Subsequently, they estimate the camera network topology and refine it based on the results of

the reidentification using the previously trained classifier. Finally, they run an online reidentification by using the inferred camera topology. In [31], each camera becomes calibrated, and they estimate the relative scales between cameras. They use calibration results to calculate the person's speed and infer the distance between cameras to generate a camera network topology based on distance. They can apply this method adaptively to each per-son according to their speed and can handle different times of transition of people between disjoint cameras.

In [32], they classify approaches based on learning models, as supervised and unsupervised. Supervised learning approaches assume the availability of a large number of manually-labeled cross-view identity matching image-pairs. They do it for each camera pair to induce a feature representation or a distance metric function optimized just for that camera pair. In [33], they propose a deep model named integration convolutional neural network (ICNN) for people ReID, which jointly learns global and local features. They extend the global features by directly performing global average pooling on the convolutional maps for each person's image. They learn the local features by performing local horizontal average pooling on the convolutional maps. They obtain several banding convolutional features for describing pedestrian parts. Finally, they concatenate global and local features by using a weighted strategy to represent the pedestrian image. In [34], they proposed it as a method for learning global and weighted local body-part features from pedestrian images. In that work, they employ angular loss and part-level classification loss jointly as a similarity measure and used for training a network robust to feature variance. Another method based on learning global and local features is in [35]. Firstly, they explore multi-feature extraction with different spatial levels projected into a shared space to reduce the dimension. After that, they use a weighted fusion approach combined dictionary learning-based sparse representation with collaborative representation. A common problem with this type of approach is the lack of discriminative features that aggregate both the spatial and temporal information. They ad-dress this problem in [36] with a joint attentive spatial-temporal feature aggregation network (JAFN), which simultaneously learns the quality- and frame-aware model to obtain attention-based spatial-temporal feature aggregation.

Unsupervised learning approaches per-camera pairwise ID labeled training data are not required in model learning [32]. In [37], they proposed that frames, where a person appears, should get divided into a set of clusters. Those subsequently match up by using a distance measure based on the naive Bayes nearest neighbor algorithm and Spearman distance. They [7] propose two solutions to fix the ReID in a closed short-term surveillance network. One is the indiscriminative patch trim strategy. The other is the multi-instance multi-label learning method. The first algorithm finds indiscriminative patches from the image of the person, while the second algorithm detects attributes of the images that help refine the final matching process.

Traditional approaches use distance metrics to calculate the degree of similarity between descriptors. In [38], they proposed an algorithm for learning a Mahalanobis distance for ReID. Such a method has two distinctive parts. They first minimize the intraclass distances to the greatest extent to obtain the best separability of the training data, by forcing intraclass distances to be zero. Secondly, they maximize the minimum margin between different classes to promote the generalization ability of the learned metric. Due to the poor quality of cameras or the extent distance from the person, the captured pedestrian videos usually suffer from low resolution, which results in the loss of useful information contained in videos. In [21], they introduce a mean distance of multi-metric subspace to address the overfitting problem, usually presented in learned metric subspace-based methods. The joint discriminant optimal model on feedback top ranks matching pairs will enhance the discrimination of matching pairs similarity. For the same problem of overfitting, in [39], they proposed a semi coupled mapping-based set-to-set distance learning (SMDL) approach.

The appearance of the person to re-identify suffers variations due to lighting, pose, rotation, noise, occlusion, scale, different cameras, among others. Various approaches came up that attempt to solve this problem. In [40], they proposed CamStyle, which serves as a data augmentation approach to smooth data disparities. They learn camera aware style transfer models from the real training data between different cameras. For each real image, they utilize the trained transfer model to generate images that fit the style of target cameras. Subsequently, real images and style-transferred images get combined to train the ReID convolutional neural network (CNN). They apply cross-entropy loss and the Label Smooth Regularization (LSR) loss to real images and style-transferred images, respectively, for reducing noise. The results on the Market-1501 and DukeMTMC-ReID data-sets show that CamStyle reduces the impact of the over-adjustment. In [41], they propose the Multi-view Common Component Discriminant Analysis (MvCCDA) to simultaneously handle the variability in the appearance of the object tracked, discrimination, and non-linearity. To achieve that, MvCCDA incorporates supervised and local geometric information into the standard component extraction process to learn a common discriminant subspace. It is useful to discover the nonlinear structure embedded in multi-view data. The method yielded promising results on databases of hand-written digits, faces, and other objects. Another approach that addresses variations in the appearance of an object followed is that proposed in [42], changes explicitly in geometry/photometry, camera point of view, illumination, and partial occlusion. The authors adopt the principles of cognitive psychology to design a flexible representation that can adapt to changes in the appearance of the object during the tracking.

Most of the previously described works require several images to re-identify the person of interest. However, there are some approaches that try to re-identify the person using a single image. In [43], there is an algorithm for learning a

Mahalanobis distance for person reidentification when only a single image exists per person. This method obtains the best separability of the training data and promotes the generalization ability of the learned metric by maximizing the minimum margin between different classes.

The method proposed in this paper use face images extracted for videos captured in a camera network to construct a descriptor for the person of interest. Because camera network topology inference is beyond the scope of this research work, we manually annotated the topologies used here. However, the proposed method can complement and extends the capacity of the approaches for cam-era network topology inference previously described. For more details about requirements for person reidentification, refer to [1], [2], [4], [44].

III. METHOD FOR PREDICTING THE TARGET FIELD-OF-VIEW IN THE REIDENTIFICATION PROBLEM

For a better understanding, some terms are defined before to present the proposed method:

- Field of view: real-world extension that can be observed at any time by a camera.
- Target field of view: field of view that the person chooses to visit.
- Path: it is a sequence of fields of view that a person visits.
- Facial descriptor: an single signal created by synthesizing a set of facial images belongs to the same person.
- Matching: operation that involves comparing two signals to determine how closely they resemble.
- Pheromone evaporation: decrease in pheromone intensity over time because of evaporation.
- Pheromone accumulation: increase in pheromone intensity over time.

As mentioned above in Section I, semi-real time ReID is computationally expensive. Computational cost may decrease by predicting the target FoV where a person could appear after leaving another FoV. That is the goal of the method described in this section. It requires that a camera network mapped onto a graph be implemented as an adjacency list or adjacency matrix.

A camera network can be modeled as a graph $G(B, A)$, in which the vertices, B, correspond to fields of view, and the edges, A, represent the relationships among the cameras (see Figure 1). The possibility that a person in the FoV i has to go to the FoV j induces these relationships. The weight over the edges corresponds to the level of preference that the person has to go from i to j. This research proposes that the weight be given by artificial pheromones, whose quantity should gradually increase as the person uses the relationship represented by the edge.

The graph of the camera network serves the proposed method of creating and initializing the pheromone table. Although the graph is constructed and initialized only once, it goes stored and updated when the network changes due to the addition or removal of cameras. The updated graph allows the restructuring of the pheromone table. This makes the proposed method invariant to changes in the size of the camera network.

Several techniques can be used to characterize a per-son's preference for a path, for example, frequency and deep learning. Through frequency, we can keep track of how many times a person visits the same FoV; in this way, the FoV most frequently in a neighborhood can be selected as a target one. Through frequency, however, it is not easy to represent the selection of a new target FoV for a known path. Besides, it is not easy to implement the gradual reduction of pheromones to emulate the forgetting of unvisited FoVs. On the other hand, one can train a deep learning model with paths previously traveled by the user. However, deep learning requires thousands of pieces of information to achieve acceptable performance. That is not practical in real biometric applications where sometimes only one sample is available. Besides, adding new paths requires a computationally expensive re-training process. For those reasons, there is a need for a simple approach that efficiently characterizes a person's preference for a path in a camera network. An approach that requires few data and can learn new target FoV and forget those no longer used.

The proposed method for predicting the target FoV is composed of two phases. The first phase consists of training a prediction model from a set of paths and videos of the person of interest. The training phase is given by Algorithm 1, where the input is a set of paths R, a sequence of images, S, associated with R, a graph, G, representing the camera network, and the identity, ID, of the person. For each path $R_i \in R$, the face of the person is detected and tracked across disjoint fields of view. The collected data are used to initialize a pheromone matrix, F, and the memory of paths, m, (See Figure 3) for a specific person whose identity is given by ID. The vector, m_{ID}, is updated only with paths that have not been previously

Algorithm 1 Initialization of Prediction Models

Data: R, S, G, ID
Result: F_{ID}, m_{ID}, bt_{ID}
initialization F_{ID} on G;
for $r \in R$ **do**
  for $i \in r$ **do**
    Detect face in the video $s \in S$ associated with the FoV i;
    Track face on the video $s \in S$ associated with i;
    Build the descriptor $H_j(k, l)$ using face images obtained in the tracking;
    Match the descriptors $H_j(k, l)$ and $H_{j-1}(k, l)$;
    Update F_{ID};
  end
  Update m_{ID} with R_i;
end
Build the biometric template bt_{ID} with the most representative face images;

Algorithm 2 Facial Reidentification With Target FoV Prediction

```
Data: G
Detect face in a FoV i;
Obtain the identity ID of the detected face;
Load F_ID and m_ID for the identified person;
Track the detected face in FoV i;
Build the facial descriptor H_i(k, l) using face images obtained in the tracking;
while person remains in G do
    Obtain an ordered list L of neighbors of the FoV i;
    while the person is not found do
        Detect face in the FoV j ∈ L;
        Track the detected face in the FoV j ∈ L;
        Build the facial descriptor H_j(k, l) using face images obtained in the tracking;
        Match the facial descriptors H_i(k, l) and H_j(k, l);
        if person is found then
            i = j;
            H_i(k, l) = H_j(k, l)
        end
    end
    Update F_ID considering the result of the prediction;
end
Update m_ID with the path followed by the person in G;
Apply evaporation to F_ID;
```

saved. If the path followed by the person is already in m_{ID}, then only the quantity of pheromones in the F_{ID} is increased. The biometric templates bt for the person of interest gets built using a representative set of face images detected in S, captured during the training process.

The pheromone matrix F and the memory of paths m created in the training process are used by Algorithm 2 each time that a person of interest visits the camera network. First, a face gets detected in any FoV that acts as an input to the area observed by a camera network. Next, the detected face is recognized for obtaining the identity ID of the person. This ID is useful for loading the data structures F_{ID} and m_{ID} for the identified person. Then, the person is tracked in the FoV by their face image, and the facial descriptor $H(k, l)$ is created once the person disappears from the camera view. Next, the reidentification is initiated according to Algorithm 2. While the person remains in the camera network represented by G, an ordered list L of neighboring FoVs is obtained each time that the person leaves an FoV. The ordering is in descending order according to the amount of artificial pheromones on the edge.

The first FoV in L has the highest probability of being visited by the person. Thus, starting from the first FoV in L, each of them gets analyzed in search of the person. If the person gets found, then F_{ID} is updated. This process repeats until the person leaves the camera network. Finally, once the person leaves the camera network, m_{ID} gets only updated if the path followed by the person has not become previously registered. Beside, evaporation is applied to F_{ID} to delay faster convergence and favor the exploration of different paths during the entire ReID process. Both the memory of paths and evaporation help the proposed method to escape from cycles in G.. Below, there is a detailed description of the target FoV prediction. It includes stagnation prevention, reduction in failed predictions, construction of facial descriptors, and identification.

A. TARGET FIELD OF VIEW PREDICTION

Figure 2 shows the general scheme of the correlation filter-based reidentification algorithm implemented in this work. Given a video sequence captured in an FoV, the reidentification works as follows. First, a face image gets detected in a video frame when the person is seen in the camera network for the first time. Next, the detected face gets tracked while the person remains visible to a camera. Then, one selects the most representatives face-images captured in the tracking to synthesize a correlation filter to use as the descriptor for the person. Finally, when the person leaves the FoV i, the target FoV j that the person could visit must be predicted immediately.

The person of interest chooses the target FoV, j, based either on the shortest distance or the configuration of their environment. This selection indicates that the person has different levels of preference for each FoV in a neighborhood V. This level of preference could be modeled, as previously mentioned, using artificial pheromones from the ACO algorithms. For implementation purposes, the amount of pheromones, $\tau_{i,j}$, among the source FoV, i, and the target FoV, j, becomes accumulated in a pheromone matrix, F, as in [26]. Moreover. the different paths that the person takes get stored in a vector m, (see Figure 3).

A pheromone matrix F becomes initialized for each person known by the system as follows. If there is a relationship between i and j, then $F(i, j)$ is initialized with the value $\tau_{i,j} = \frac{1}{|V|}$, where V is the set of FoVs that are neighbors of i. Otherwise, it initializes with $\tau_{i,j} = 0$. The following matrix is an example of the result obtained by this initialization process on the graph in Figure 1:

$$F = \begin{bmatrix} & f & i & j & k & l \\ f & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \\ i & 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \\ j & 0.25 & 0.25 & 0.00 & 0.25 & 0.25 \\ k & 0.33 & 0.33 & 0.33 & 0.00 & 0.00 \\ l & 0.33 & 0.33 & 0.33 & 0.00 & 0.00 \end{bmatrix}$$

Each FoV $j \in V$ has a probability P of being visited by the person that leaves i, which is given by:

$$P(j) = \frac{\tau_{i,j}}{\sum_{l \in V_i}^{|V_i|} \tau_{i,l}}, \tag{1}$$

where l is a neighbor of i. One can get the ordered list L by performing this equation for each FoV in the current neighborhood, inserting the resulting value into L in

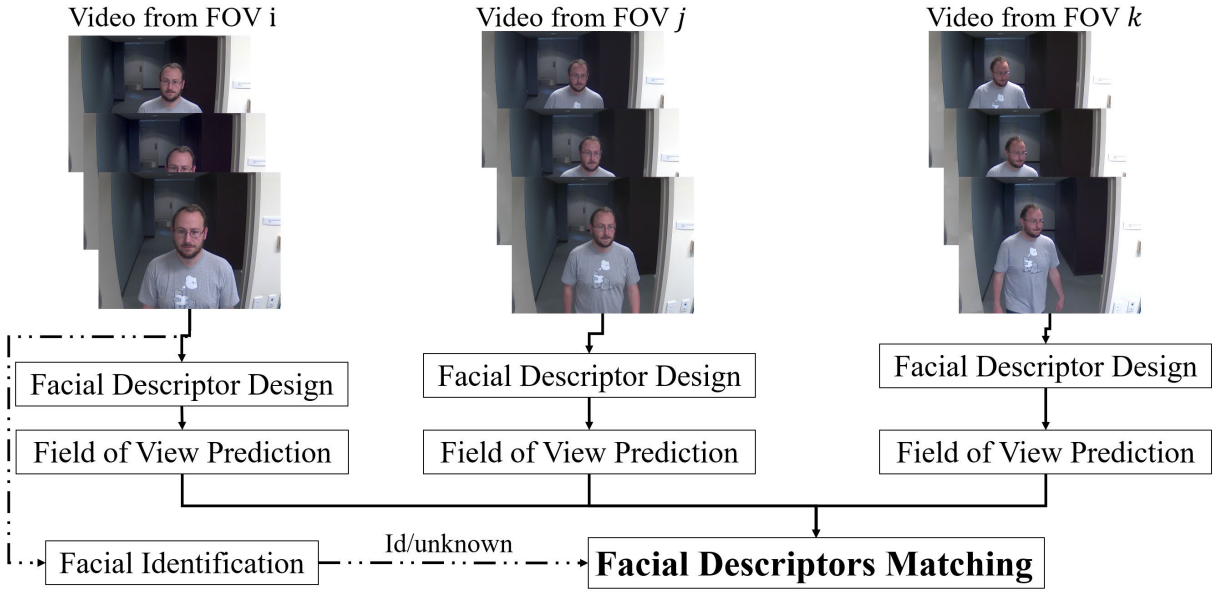

FIGURE 2. Basic scheme of video-based person reidentification using correlation filters.

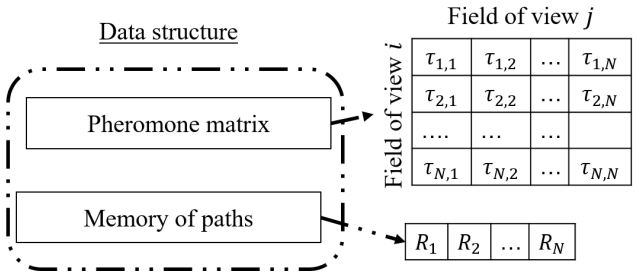

FIGURE 3. Data structures used by the proposed method: pheromone matrix F (above) and memory of paths m (below).

descending order. The FoV in the first place in L has the highest probability to be visited, and therefore, it becomes chosen as the target FoV. When the person arrives at the target FoV, j, the level of pheromones $\tau_{i,j}$ in F gets updated with:

$$\tau_{i,j} = \tau_{i,j} + \frac{\tau_{i,j}}{\sum_{l \in V_i}^{|V_i|} \tau_{i,l}}. \tag{2}$$

This updating is only performed if the predicted target FoV and the FoV visited by the person are the same, which means a successful prediction. Assume that a person is in the FoV i, of the graph in Figure 1, with pheromone matrix previously presented. The person has three possible destinations: j, k, or l; according to Equation 1, each of these FoVs has a probability of 0.33 of being visited. In this case, we suggested that the target FoV appears in the first position in the list L. Thus, FoV j, must get selected as the target, and the matrix F is updated with $\tau(i, j) = 0.33 + 0.33 = 0.66$, according to Equation 2.

B. PHEROMONE EVAPORATION

The continuous deposit of pheromone over an edge can cause excessive accumulation and cause stagnation of the method. It is necessary to undertake evaporation of pheromones in F when the person leaves the camera network to avoid this problem:

$$F = \left(1 - \rho^{-r}\right) F, \tag{3}$$

where ρ is the number of FoVs where the person got viewed, and r is the number of times that the person has visited the camera network. The factor r, is the rate of evaporation.

C. MINIMIZING FAILED PREDICTIONS CAUSED BY CYCLES

The presence of cycles in G can provoke an excessive accumulation of pheromones, causing failed prediction; however, faults can also become evident because the person has selected a new target FoV. In this work, one used two strategies adapted from those described in [26] to minimize the effect of this problem:

- *Gradual reduction of number of artificial pheromones to accumulate*. The idea here is to reduce the number of pheromones deposited every time the person visits the same FoV. To achieve this, the modification of the denominator of Equation 2 is as follows:

$$\tau_{i,j} = \tau_{i,j} + \frac{\tau_{i,j}}{\left(\sum_{l \in V_i}^{|V_i|} \tau_{i,l}\right)^q}, \tag{4}$$

where q is the number of times that the person has visited j during the current tour. This process gradually reduces the level of pheromones to be deposited. The first time that the person visits j, the level of pheromones deposited

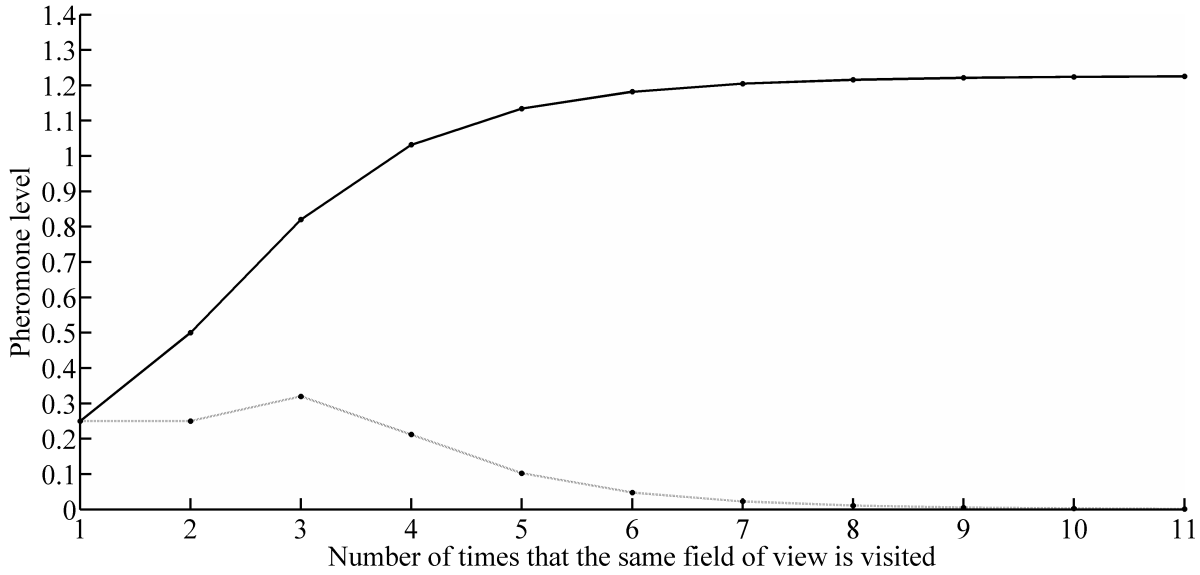

FIGURE 4. Level of pheromones on the path of the person in each visit to the same FoV (continuous line) and the amount of pheromone deposited in each visit (dotted line).

is equal to $\frac{1}{|V_i|}$. During subsequent visits to j, the quantity of pheromones deposited can vary and tend to gradually decrease, as shown by the dotted line in Fig. 4. The continuous line shows the increase in the level of pheromones every time that the person visits j until reaching stability.

- *Penalization of failed predictions*. This approach seeks to apply a reduction in pheromones presented in the relationships that cause a failed prediction. This penalization is given by:

$$t_{i,j} = (1 - \beta)\,\tau(i, e), \tag{5}$$

where $0 < \beta < 1$ is the penalization factor that acts as an escape mechanism for promoting the breaking of the relationship that causes the failed prediction. The penalty is applied to the FoV j because the person has decided to change to a new target FoV. Thus, the FoV j must become gradually forgotten.

In Algorithm 2, the update of F for fields of view, i, and the selected by the persons is performed by using equations in Eq. 2 and Eq. 4. Furthermore, F is updated according to Eq. 5 in the case of a failed prediction. This single-use of these equations allows the proposed method to forget a not-used target FoV and to learn a new target FoV.

D. CORRELATION FILTERS AS FACIAL DESCRIPTOR

Correlation filters (CF) are used in pattern recognition due to following advantages [20], [27], [45], [46]: a) invariant to noise, shift, and variable illumination, b) high ability for discrimination, c) ability to use both content and shape, d) can simultaneously detect and locate, and e) ability to continually adapt to changes in appearance. Correlation filters can be applied to face detection [47], [48], tracking [49]–[51] and recognition [52].

Let $C = \{f_1(x, y), \ldots, f_N(x, y)\}$ be the set of N facial images captured during facial tracking in FoV i. Using all face images, including images of poor quality, can degrade the quality of the correlation filter. Thus, it is advisable to select only the best subset $T \subset C$ for training a CF. This selection can get performed using the approaches proposed by [52] or [53]. The correlation filter used as a descriptor in this work is given by [50]:

$$H_i(u, v) = \frac{1}{N} \sum_{n=1}^{N} F_n^k(u, v), \tag{6}$$

where $F_n(u, v)$ is the Fourier Transform (FT) of $f_n(x, y) \in T$, and $0 < k < 1$ is the nonlinear factor [54]. This correlation filter was selected as a biometric template because it can recognize incomplete face images. Such images are common under unconstrained environments due to occlusion, pose, or point of view. The descriptor, $H_i(u, v)$, is correlated against the descriptor $H_j(u, v)$ via a correlation process [27]:

$$g(x, y) = \mathscr{F}^{-1}\{H_i(u, v)H_j^*(u, v)\}, \tag{7}$$

where $\mathscr{F}^{-1}$ is the inverse of the FT (IFT), and $H_j(u, v)$ is the facial descriptor generated in target FoV j. The correlation output $g(x, y)$ is examined toward searching a correlation peak whose sharpness is characterized by the measure known as peak-to-sidelobe ratio (PSR) [27]. If $PSR \geq \tau$, then this means that the descriptor $H_j(u, v)$ resembles the descriptor $H_i(u, v)$, and there is a correspondence.

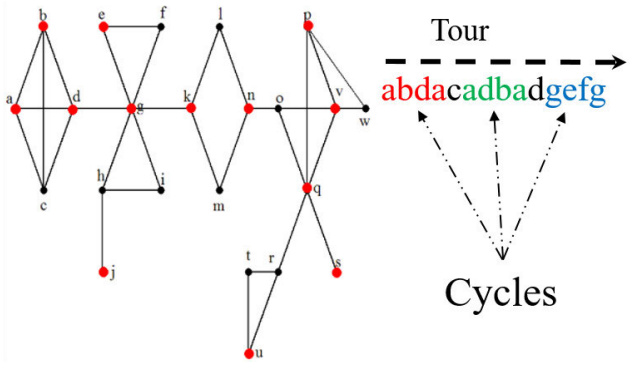

FIGURE 5. Graph depicting a camera network with the entrances and exits shown in red (left) and a path followed by a person on tour (right).

E. FACIAL IDENTIFICATION

Reidentification algorithms are not required to know the *ID* of the person of interest, which can remain unknown during the entire process. However, knowing the person's identity allows reidentification algorithms to work under biometric contexts. Because of this, the proposed approach includes a recognition module that can be performed only once in any part of the ReID process. Facial recognition works as follows. Each descriptor gets compared against all the descriptors stored in the gallery. If there is a match, then the identity *ID* of the person gets reported. The gallery is a database containing descriptors generated by the Eq. III-D.

F. PREDICTION RATE

The performance of the proposed method is measured by calculating the prediction rate (PR), which is given by:

$$PR = \frac{Number\ of\ successful\ predictions}{Total\ of\ predictions}. \tag{8}$$

A successful prediction occurs when the FoV selected by the person is the same as that predicted by the proposed algorithm.

This section presented an algorithm to predict the target FoV, which uses correlation filters as the person's descriptors. The next section describes the results obtained in an experimental evaluation.

IV. EXPERIMENTAL RESULTS

The results obtained from two the experiments are presented in this section. First, the proposed method was used on 31 simulated scenarios and two real scenarios to test their ability to predict the target FoV. Second, we present an analysis of the expected path and the path followed by the person. We manually annotated paths used in training, while in the test, the paths were defined by the trajectory that the person followed in the camera network. The prediction rate avoids the need to annotate the test paths manually; that is because the path followed by the person is compared directly against the information of paths stored in the pheromone matrix.

Before these two experiments, we computed the performances of approaches for minimizing the effect of failed prediction presented in Sec. III-C. We performed four experiments on the graph depicted in Fig. 5: 1) updating the pheromones via Eqs. 4 and 5; 2) updating the pheromones via Eq. 2 and Eq. 5; 3) updating the pheromones via Eq. 4; and 4) updating the pheromones via Eq. 2. For each experiment, thirty artificial ants were used to emulate people, with each one performing 500 tours on the camera network.

Figure 6 shows the performance obtained with the first and second experiment. Penalizing failed predictions achieved the best performance using $\beta = 0.9$,

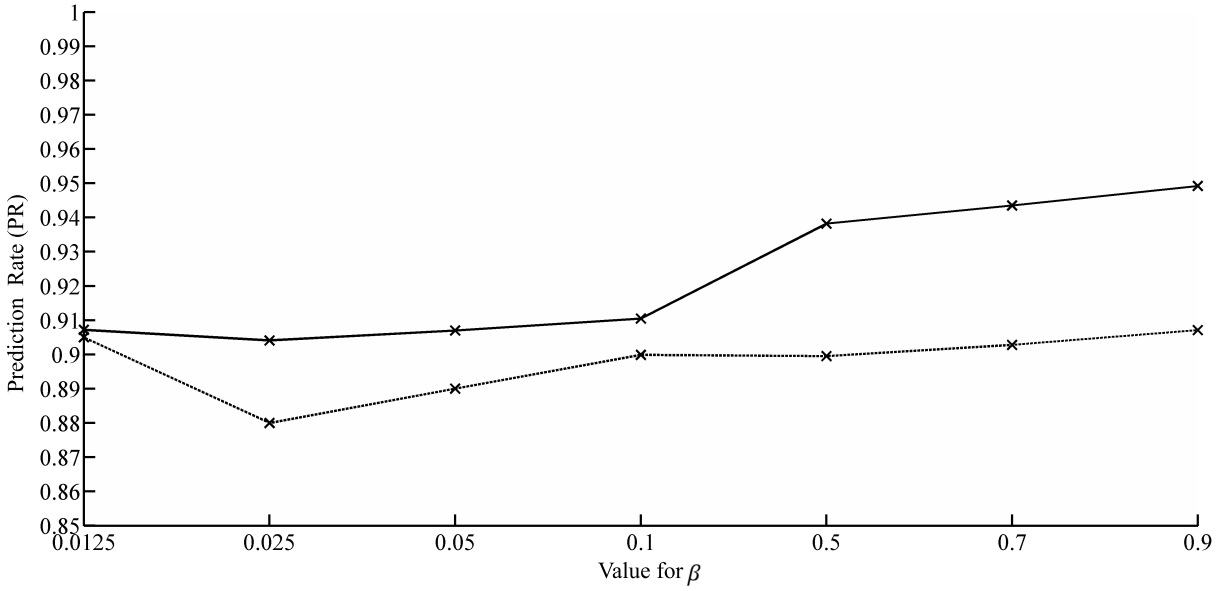

FIGURE 6. Performance in terms of the average number of cycles for experiment 1 (dotted line) and experiment 2 (continuous line) with different values for β.

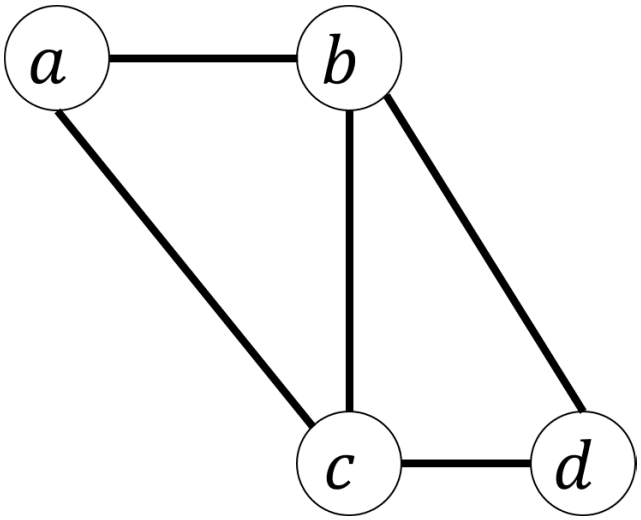

FIGURE 7. Graph of the camera network where the LPA2 dataset was obtained.

outperforming the results of third and fourth experiment that obtained *PR* performances equal to 0.8778 and 0.9140, respectively. Updating the pheromone and penalizing failed predictions (second experiment), an average of 574 cycles occurred compared to 709 when no penalty was applied (fourth experiment). The use of factor q did not reduce the occurrence of cycles in the third experiment, which was a consistent trend because this only causes a slow artificial pheromone accumulation.

The results presented in the rest of this paper were obtained by updating the level of pheromones via Eq. 4 and penalizing the failed predictions via Eq. 5. It is important to note that the penalty of failed predictions helps the proposed method to forget unexploited targets. It also improves the capacity to learn a new target FoV without a relearning process.

A. EVALUATION ON A REAL SCENARIO

Consider the following assumptions in real scenarios. First, the person has the ability to detect and leave any cycle in the camera network at any time. The success of the prediction depends on the quality of face images collected by the detection and tracking modules and the robustness of the descriptor to changes in the appearance of the person.

TABLE 1. Datasets from real scenarios used in the evaluation of the proposed method.

Dataset	Subjects	FoVs	Variation
ChokePoint	24	4	Illumination, pose, sharpness, and misalignment.
LPA2	1	4	Illumination, scale, pose, and scale.

1) DATA CONFIGURATION

We considered real scenarios in this evaluation. Firstly, the publicly available ChokePoint [55] video dataset designed for people reidentification under real-world surveillance conditions. We selected sequences of 24 persons from this dataset and organized such that each person crosses four fields of view. Faces have variations in terms of illumination, pose, sharpness, and misalignment due to automatic face localization/detection. The videos got recorded in two portals with a month of difference between them; this is the agreement with the creators of the data set. LPA2 is the name of the second dataset used in this research. The videos became captured from a camera network in a laboratory with only one access door and one wall that divides the laboratory into two sections. The cameras were placed to capture the facial image of the person when entering and leaving in each section. This scenario contains different light sources with different intensities and receives exterior light through the windows and the door. Figure 7 presents this scenario, comprising fields of view a, b, c and d, in which the camera network is entered through the FoV a, while the exit is through FoV b.

FIGURE 8. Examples of frames in the LPA2 (first row) and ChokePoint (second row) datasets.

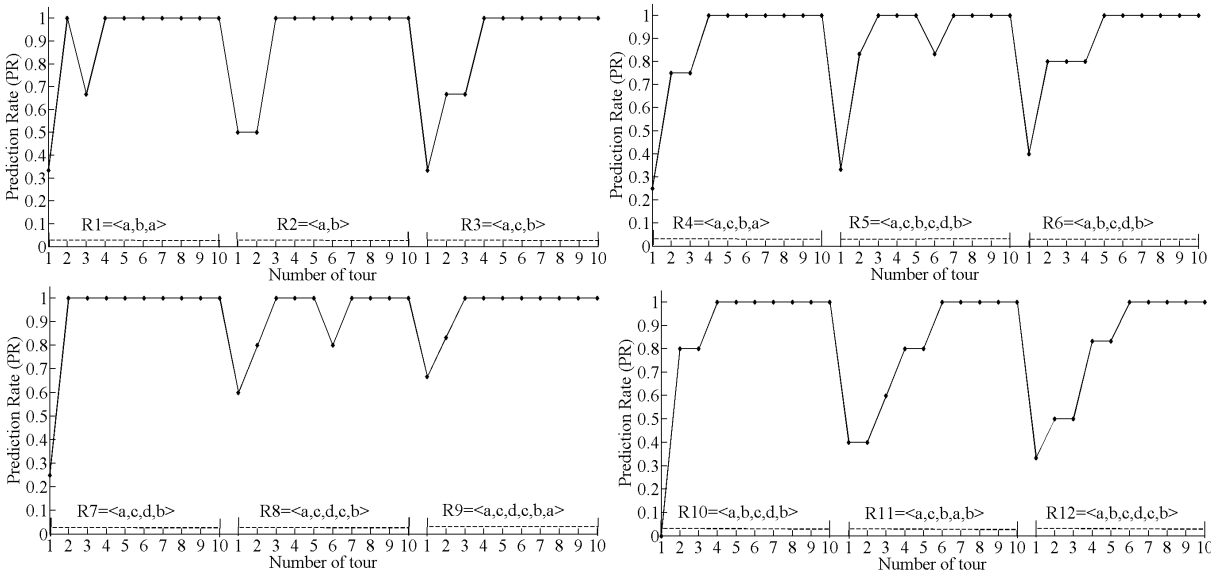

FIGURE 9. Results on the LPA2 data set with twelve routes sequentially traveled, where the person toured each of them ten times.

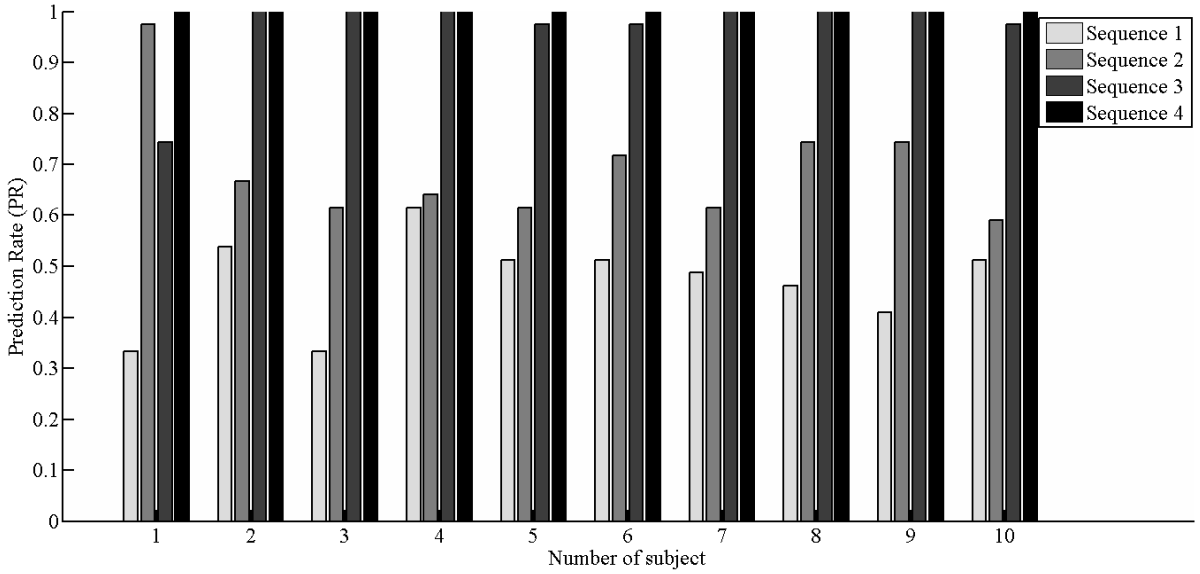

FIGURE 10. Results on a subset of ten people from the ChokePoint dataset with thirteen executions. Four sequences were used for each person, where each sequence consists of four videos.

TABLE 2. Performance of the proposed method in terms of the Levenshtein distance.

Path	Subject																							
	0003	0004	0005	0006	0007	0009	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	0020	0021	0022	0023	0024	0025	0026	0027
R_5	0	0	0	0	0	0	0	0	0	5	0	0	1	1	0	0	0	0	0	1	0	0	0	0
R_6	0	0	4	0	2	0	0	1	0	0	0	0	4	0	0	5	0	0	2	3	0	3	0	5
R_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R_8	0	0	0	0	0	0	0	0	4	4	0	0	0	0	4	0	0	0	0	0	0	0	0	0
R_9	4	0	2	1	2	0	3	0	0	1	0	0	0	0	0	0	0	3	0	0	0	2	0	0
R_{10}	0	0	0	0	0	0	0	0	0	0	0	4	0	0	4	0	0	0	0	3	0	0	0	0
R_{11}	0	0	0	0	4	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
R_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0

Figure 8 shows some frame samples of the LPA2 dataset (first row) and the ChokePoint dataset (second row).

ChokePoint and LPA2 datasets are summarized in Table 1. ChokePoint contains 24 people, four FoVs, and facial variation under indoor/outdoor conditions. LPA2 dataset was created for this project, and it is not public. It contains videos of one person, a network of four cameras and the face presents variations due to indoor conditions.

FIGURE 11. Frame samples with correctly identified persons in the dataset extracted from ChokePoint.

TABLE 3. Results on the paths R_{5-12} in the Figure 13 and the ChokePoint dataset.

Path	Subject																							
	0003	**0004**	**0005**	**0006**	**0007**	**0009**	**0010**	**0011**	**0012**	**0013**	**0014**	**0015**	**0016**	**0017**	**0018**	**0019**	**0020**	**0021**	**0022**	**0023**	**0024**	**0025**	**0026**	**0027**
R_5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.80	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	1.00
R_6	1.00	1.00	0.20	1.00	0.60	1.00	1.00	0.80	1.00	1.00	1.00	1.00	0.20	1.00	1.00	1.00	1.00	1.00	0.60	0.20	1.00	0.20	1.00	1.00
R_7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R_8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R_9	0.00	1.00	0.00	0.67	0.00	1.00	0.00	0.67	1.00	0.00	1.00	0.67	1.00	1.00	0.67	1.00	1.00	0.67	1.00	0.67	1.00	0.00	1.00	1.00
R_{10}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.25	1.00	1.00	1.00	1.00
R_{11}	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R_{12}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Average	**0.88**	**1.00**	**0.78**	**0.96**	**0.70**	**1.00**	**0.88**	**0.93**	**1.00**	**0.88**	**1.00**	**0.96**	**0.88**	**0.98**	**0.96**	**1.00**	**1.00**	**0.83**	**0.95**	**0.74**	**1.00**	**0.78**	**1.00**	**1.00**

2) RESULTS ON REAL SCENARIOS

Experiments in this section used the complete ReID system, which includes the tasks of detection, tracking, identification, and the proposed method to predict the target FoV. The face images got detected by the approach proposed in [56], while for tracking, the correlation filters-based algorithm described in [50].

We evaluated twelve paths on the LPA2 dataset. On three of these paths, video sequences were recorded registering a person who used each path ten times. These videos had

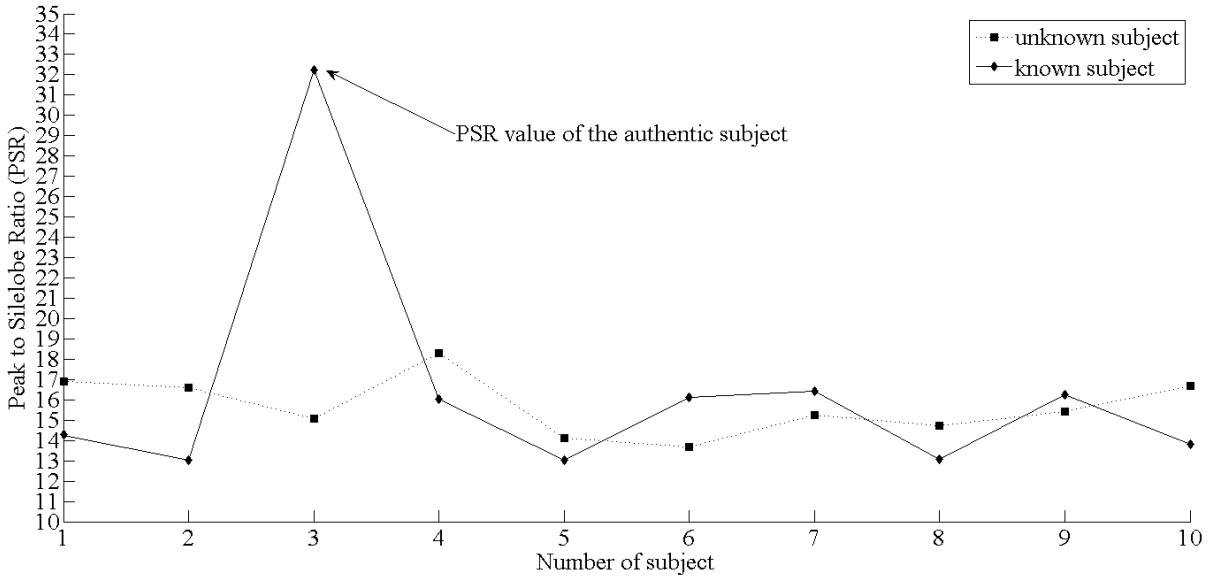

FIGURE 12. PSR performance in the identification of a person with a biometric template (person number 3) and a person without a biometric template (unknown person).

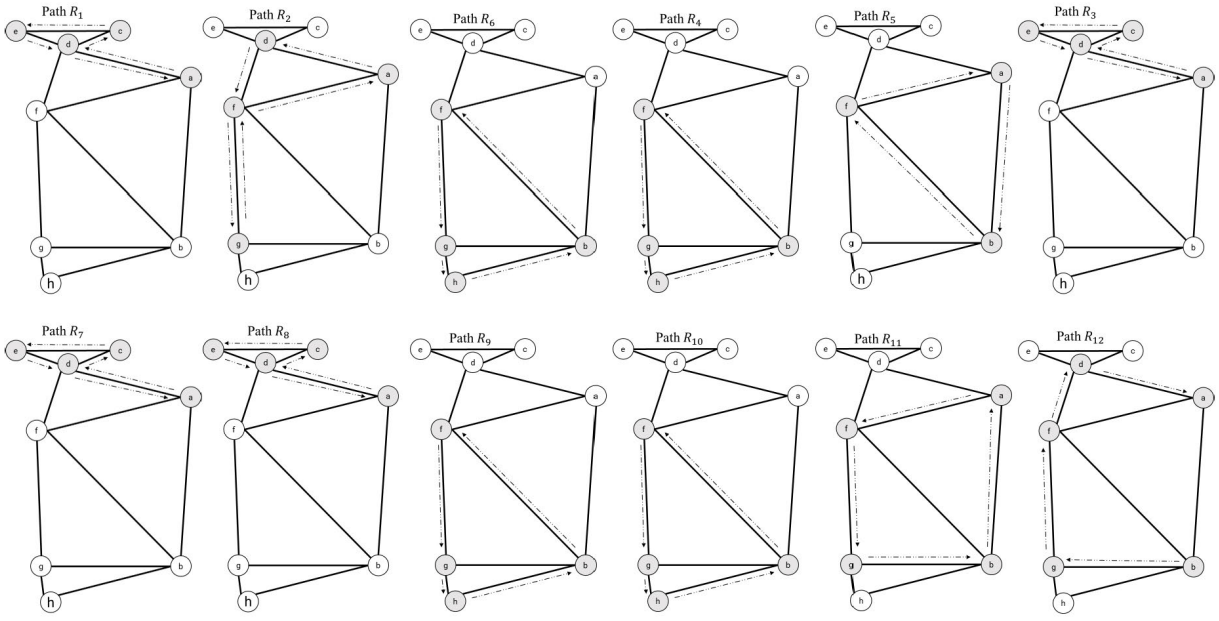

FIGURE 13. Paths on a camera network used in the evaluation with 24 people from the ChokePoint dataset. Paths R_1, R_2, R_3, and R_4 became used for training, while the rest of the pats became used for test the proposed method.

a rate of 30 frames per second, a resolution of 480 × 640 pixels, and grayscale images. The appearance of face images is affected by illumination, pose, rotation (in-plane and out-plane), scale, partial occlusion, and facial expressions. On these three paths, the complete reidentification system was tested, including face detection, tracking, and the synthesis of the facial descriptor. The remaining nine paths were manually generated to analyze the performance of the prediction model in a real scenario in isolation from the rest of the reidentification system.

Figure 9 shows the *PR* performance of the proposed algorithm on the LPA2 dataset. Twelve paths were used for the evaluation, and the person used each of them ten times. The proposed method obtained an average $PR = 0.8256$ and required that the person visited at least four times the same path to reach a $PR = 1.0$. The *PR* performance decreases when a new target FoV is selected by the person. However, the performance gradually increases as the proposed method adapts to the new target FoV. Paths R_{11} and R_{12} were challenging because they presented several cycles and long paths.

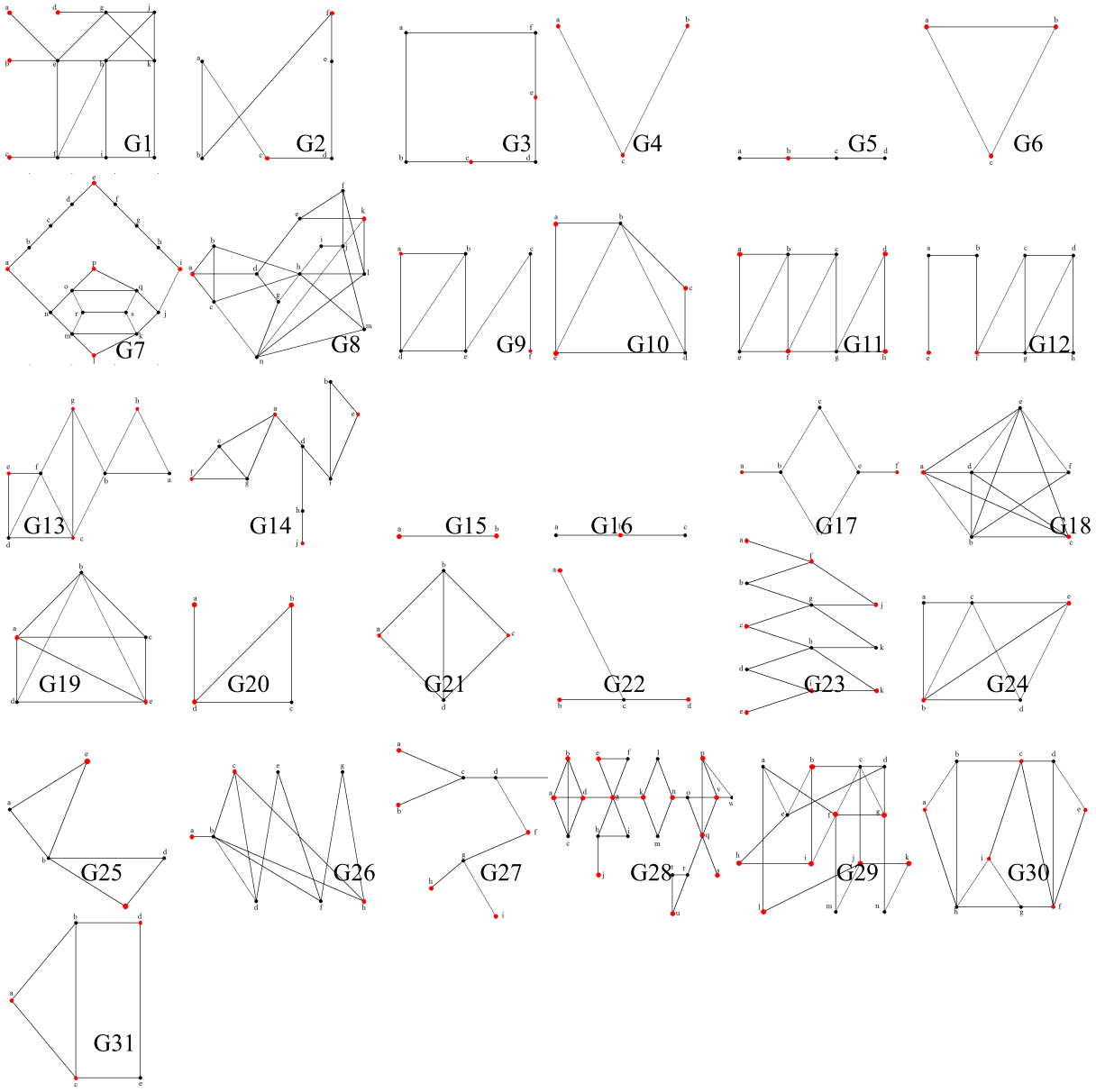

FIGURE 14. Graphs of the camera networks in the simulated scenarios.

Figure 10 depicts the average *PR* performance obtained on the ChokePoint dataset with 13 repetitions of the experiment, ten people, and four sequences. In most cases, performance improved proportionally to the number of paths traveled by the person. Only the first person obtained a decrease in performance from second and third sequence; this because, in the third sequence, there were frames where the face images contained shadows and intense illumination, causing mismatching between descriptors. The proposed method obtained an average *PR* equal to 0.7850 along with the sequences and subjects, and it reached an average of 1.0 in the last sequence for all subjects.

An average of 40 facial images got captured in each video in LPA2, while for ChokePoint, an average of 34 got captured. From these sets, the best subsets became automatically selected by the method in [52] for synthesizing the descriptor of persons that obtained recognition rates of 100%. Although there was a month between the recording of videos in portal 1 and portal 2 of ChokePoint, the correlation filters were able to perform a correct matching in the reidentification. These results were not affected by the different clothes that the person used, which is an advantage to using faces in the person reidentification.

A detection and identification rate (DIR) [57] equal to 100% in the identification of persons on tested datasets got

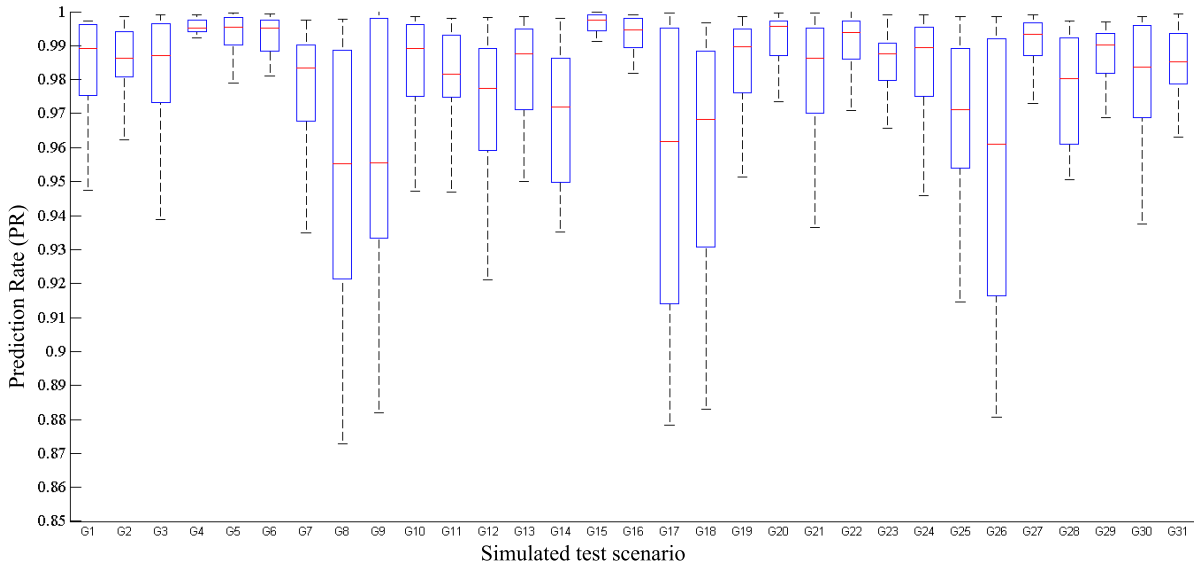

FIGURE 15. Box plot of prediction rate (*PR*) performance on the simulated scenarios.

obtained. An average of ten different facial images got used to training the biometric templates. Each facial image was grayscale with a resolution of 128 × 128 pixels. Comparing a suitable facial image captured in an FoV against the biometric templates produced a set of similarity scores. Thus, the identity of the facial image corresponds to the biometric template with the *nth* most significant *PSR* value. Figure 11 shows frame samples with correctly identified persons in the dataset extracted from ChokePoint. As shown in Figure 12, the correlation filter correctly identified a known person while rejecting unknown persons.

To get a more in-depth analysis of the expected path and the path followed by the person of interest, we performed an experiment using the 24 people of the ChokePoint dataset. We got the results presented here on the twelve paths depicted in Figure 13. The first four paths became used to train the proposed method, while the remaining eight paths became used for the tests. The paths got used sequentially by the people to test the capacity of the proposed method to produce paths similar to expected paths. Besides, we evaluated the ability of the proposed approach to using information from previously learned paths to predict the target FoV and adapt to changes in paths.

Table 2 shows the result of this evaluation in terms of the Levenshtein distance. It measures the difference between the expected path and the path followed by the person. In most cases, a distance equal to zero got obtained, which means that compared paths are similar, while a distance equal to four or five means that the person followed a different path than expected. Furthermore, it was observed in the experiment that 70% − 80% of neighbors were analyzed in search of the person of interest. This result indicates that there was a savings of 20% − 30% of computational resources and time.

Table 3 contains the performance of the proposed method, where the proposed method obtained a *PR* greater than or equal to 0.78. These results are promising and provide evidence that the proposed method can learn full paths and adapt to small changes.

B. EVALUATION ON SIMULATED SCENARIOS

We could not find a public dataset containing sequences of videos recorded on several suitable camera networks to test reidentification algorithms. Therefore, the proposed method was evaluated on simulated scenarios to observe its performance in networks with different features. We used the Monte Carlo method to measure the performance of the proposed algorithm on the simulated scenarios. People and camera networks became emulated by artificial ants and non-directed graphs, respectively. Each artificial ant performed 2000 tours on each graph. First, an FoV is randomly selected to access the camera network. Second, the Monte Carlo method is used to select the FoV that will be visited by the artificial ant. This step emulates the process that the person follows to select the place to visit. Third, by the proposed method, the target FoVs that the ant could visit becomes predicted. Fourth, the pheromone level on the trail gets updated. Steps two to four get executed iteratively until the artificial ants leave the camera network.

Figure 14 shows graphs extracted from Refs. [26] and [58], used in the simulated scenarios. The graphs are different in the number of nodes and the configuration of the links connecting them.

Figure 15 shows the box plot of *PR* performance for the proposed algorithm. We obtained an average *PR* equal to 0.9789 with a variance in 0.00059. Graphs *G*8, *G*9, *G*17, *G*18 and *G*26 obtained a performance below average and a higher

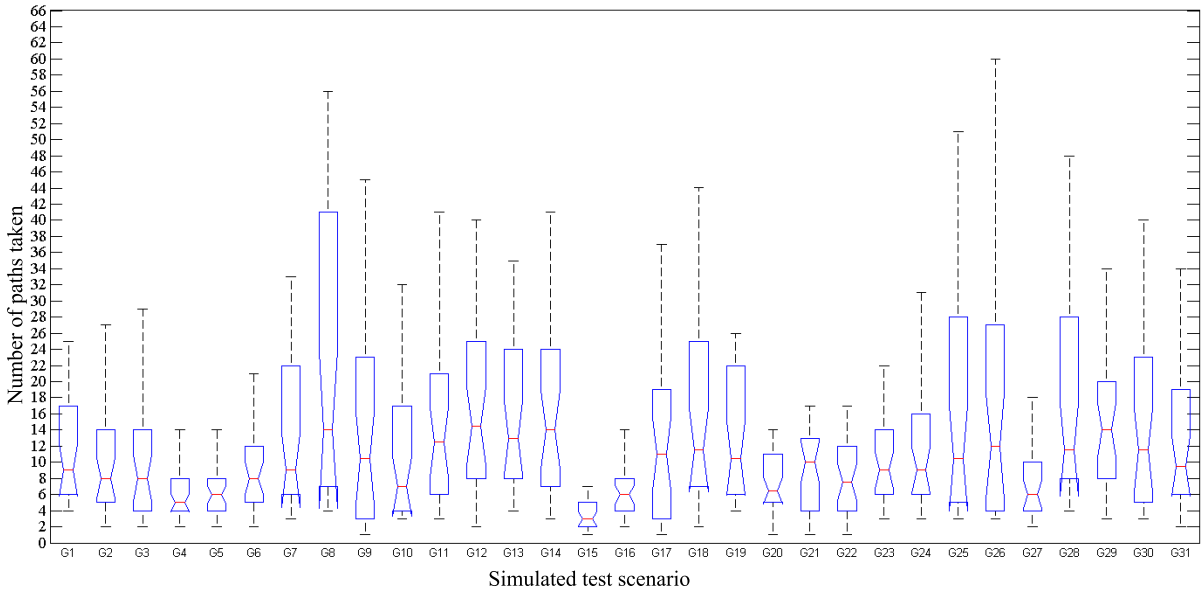

FIGURE 16. Box plot with the number of paths generated in each path on the simulated scenarios.

level of variability because they are graphs that contain various cycles, a high number of nodes, and few exits.

Figure 16 presents the box plot with the number of paths taken. On average, the artificial ants tested 16 paths before selecting the best one. Networks $G8$, $G9$, $G25$, $G26$ and $G28$ registered a more significant variation due to their higher number of cycles and fewer number of exit options.

The most significant challenges faced by the correlation filters in the facial identification were shadows and intense illumination. Preprocessing operations were applied to improve global lighting and remove the variable illumination. It was impossible, however, to recover all the border and texture information, which are vital for a person's discrimination. In facial tracking, we observed that the re-synthesis of the tracking filter in every 15 face images improves the filter's ability to learn the newest detected faces while forgetting the oldest. Additionally, this avoids obtaining over-fitting in the filter. According to the results described in this section, the proposed algorithm can predict the target FoV under real-world surveillance conditions. Furthermore, it is capable of recovering from prediction failures that can occur after some changes in the path and works both short- and long-term periods.

V. CONCLUSION

This paper proposed a method for predicting the target field of view in the facial reidentification problem based on correlation filters. The obtained results indicate the following. First, principles of the ant colony optimization algorithm can be used efficiently to indicate the level of preference that a person has for going from an FoV to a neighbor, FoV, in a scenario monitored by a camera network. Second, the traced paths show the behavior of the person of interest in a camera network and provide sufficient data for the proposed method to robustly predict the target FoV. Third, correlation filters are useful and accurate as facial descriptors for reidentification under facial variations caused by variable illumination, expressions, rotation, and scale.

As future work, we are interested in comparing our approach with other prediction meta-heuristics. Besides, we will apply the proposed method over a more significant number of real indoor and real outdoor scenarios.

REFERENCES

[1] A. Bedagkar-Gala and S.-K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, to be published.
[2] S. Gong, M. Cristani, C. Loy, and T. Hospedales, "The re-identification challenge," in *Person Re-Identification Advances in Computer Vision and Pattern Recognition*, S. Gong, M. Cristani, S. Yan, and C. C. Loy, Eds. London, U.K.: Springer, 2014, pp. 1–20.
[3] R.-C. Hario-Pribadi and H.-K. Pao, "Sparse tree structured representation for re-identification," *Pattern Recognit.*, vol. 60, pp. 394–404, Dec. 2016, doi: 10.1016/j.patcog.2016.05.003.
[4] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," 2016, *arXiv:1610.02984*. [Online]. Available: https://arxiv.org/abs/1610.02984
[5] A. Zheng, F. Wang, A. Hussain, J. Tang, and B. Jiang, "Spatial-temporal representatives selection and weighted patch descriptor for person re-identification," *Neurocomputing*, vol. 290, pp. 1–9, May 2018, doi: 10.1016/j.neucom.2018.02.039.
[6] Y. Guo, K. Zhao, X. Hao, and M. Yu, "Deep regression neural network for end-to-end person re-identification," *IEEE Access*, vol. 7, pp. 92825–92837, 2019.
[7] Y. Lin, F. Guo, L. Cao, and J. Wang, "Person re-identification based on multi-instance multi-label learning," *Neurocomputing*, vol. 217, pp. 19–26, Dec. 2016, doi: 10.1016/j.neucom.2016.04.060.
[8] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," 2017, *arXiv:1705.04724*. [Online]. Available: https://arxiv.org/abs/1705.04724
[9] M. Fischer, H. K. Ekenel, and R. Stiefelhagen, "Interactive person re-identification in TV series," in *Proc. Int. Workshop Content Based Multimedia Indexing (CBMI)*, Jun. 2010, pp. 1–6.

[10] M. Bauml, K. Bernardin, M. Fischer, H. Ekenel, and R. Stiefelhagen, "Multi-pose face recognition for person retrieval in camera networks," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug./Sep. 2010, pp. 441–447.

[11] M. D. Marsico, G. Doretto, and D. Riccio, "M-VIVIE: A multi-thread video indexer via identity extraction," *Pattern Recognit. Lett.*, vol. 33, pp. 1882–1890, Oct. 2012, doi: 10.1016/j.patrec.2012.03.005.

[12] G. Wang, F. Zheng, C. Shi, J.-H. Xue, C. Liu, and L. He, "Embedding metric learning into set-based face recognition for video surveillance," *Neurocomputing*, vol. 151, no. P3, pp. 1500–1506, 2015.

[13] Y. Wang, J. Shen, S. Petridis, and M. Pantic, "A real-time and unsupervised face re-identification system for human-robot interaction," *Pattern Recognit. Lett.*, to be published. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865518301296

[14] G. Berdugo, O. Soceanu, Y. Moshe, D. Rudoy, and I. Dvir, "Object reidentification in real world scenarios across multiple non-overlapping cameras," in *Eur. Signal Process. Conf.*, 2010, pp. 1806–1810.

[15] R. Mazzon, S. F. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognit. Lett.*, vol. 33, no. 14, pp. 1828–1837, Oct. 2012, doi: 10.1016/j.patrec.2012.02.014.

[16] T. Zhou, M. Qi, J. Jiang, X. Wang, S. Hao, and Y. Jin, "Person re-identification based on nonlinear ranking with difference vectors," *Inf. Sci.*, vol. 279, pp. 604–614, Sep. 2014, doi: 10.1016/j.ins.2014.04.014.

[17] L. Nanni, M. Munaro, S. Ghidoni, E. Menegatti, and S. Brahnam, "Ensemble of different approaches for a reliable person re-identification system," *Appl. Comput. Informat.*, to be published.

[18] L. Ren, J. Lu, J. Feng, and J. Zhou, "Multi-modal uniform deep learning for RGB-D person re-identification," *Pattern Recognit.*, vol. 72, pp. 446–457, Dec. 2017.

[19] Q. Zhou, S. Zheng, H. Ling, H. Su, and S. Wu, "Joint dictionary and metric learning for person re-identification," *Pattern Recognit.*, vol. 72, pp. 196–206, Dec. 2017, doi: 10.1016/j.patcog.2017.06.026.

[20] Q. Wang, A. Alfalou, and C. Brosseau, "New perspectives in face correlation research: A tutorial," *Adv. Opt. Photon.*, to be published.

[21] Y. Ren, X. Li, and X. Lu, "Feedback mechanism based iterative metric learning for person re-identification," *Pattern Recognit.*, vol. 75, pp. 1339–1351, Mar. 2018.

[22] A. Roy, S. Sural, and J. Mukherjee, "A hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification," *Pattern Recognit. Lett.*, vol. 33, pp. 1891–1901, Feb. 2012.

[23] Z. Liu, Z. Zhang, Q. Wu, and Y. Wang, "Enhancing person re-identification by integrating gait biometric," in *Computer Vision—ACCV 2014 Workshops* (Lecture Notes in Computer Science), vol. 9008, C. Jawahar and S. Shan, Eds. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-16628-5_3.

[24] A. Utsumi and N. Tetsutani, "Human tracking using multiple-camera-based head appearance modeling," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 657–662.

[25] G. Watson and A. Bhalerao, "Person reidentification using deep foreground appearance modeling," *J. Electron. Imag.*, vol. 27, p. 27, 2018, doi: 10.1117/1.JEI.27.5.051215.

[26] M. Dorigo and T. Stutzle, *Ant Colony Optimization*, M. Dorigo and T. Stutzle, Eds. Cambridge, MA, USA: Massachusetts Institute of Technology, 2004.

[27] B. Vijaya-Kumar, A. Mahalanobis, and R. Juday, *Correlation Pattern Recognition*, 1st ed., B. Vijaya-Kumar, A. Mahalanobis, and R. Juday, Eds. New York, NY, USA: Cambridge Univ. Press, 2005.

[28] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016786551200219X

[29] Z. Han, S. Li, C. Cui, H. Song, Y. Kong, and F. Qin, "Camera planning for area surveillance: A new method for coverage inference and optimization using location-based service data," *Comput., Environ. Urban Syst.*, vol. 78, Nov. 2019, Art. no. 101396. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0198971519300936, doi: 10.1016/j.compenvurbsys.2019.101396.

[30] Y.-J. Cho, S.-A. Kim, J.-H. Park, K. Lee, and K.-J. Yoon, "Joint person re-identification and camera network topology inference in multiple cameras," *Comput. Vis. Image Understand.*, vol. 180, pp. 34–46, Mar. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314219300037

[31] Y.-J. Cho and K.-J. Yoon, "Distance-based camera network topology inference for person re-identification," *Pattern Recognit. Lett.*, vol. 125, pp. 220–227, Jul. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016786551830730X, doi: 10.1016/j.patrec.2019.04.009.

[32] M. Li, X. Zhu, and S. Gong, "Unsupervised person re-identification by deep learning tracklet association," *CoRR*, vol. abs/1809.02874, 2018. [Online]. Available: http://arxiv.org/abs/1809.02874

[33] Z. Zhang, T. Si, and S. Liu, "Integration convolutional neural network for person re-identification in camera networks," *IEEE Access*, vol. 6, pp. 36887–36896, 2018.

[34] Y. Wang, Z. Wang, W. Jia, X. He, and M. Jiang, "Joint learning of body and part representation for person re-identification," *IEEE Access*, vol. 6, pp. 44199–44210, 2018.

[35] J. Guo, Y. Zhang, Z. Huang, and W. Qiu, "Person re-identification by weighted integration of sparse and collaborative representation," *IEEE Access*, vol. 5, pp. 21632–21639, 2017.

[36] L. Chen, H. Yang, and Z. Gao, "Joint attentive spatial-temporal feature aggregation for video-based person re-identification," *IEEE Access*, vol. 7, pp. 41230–41240, 2019.

[37] C. Riachy, F. Khelifi, and A. Bouridane, "Video-based person re-identification using unsupervised tracklet matching," *IEEE Access*, vol. 7, pp. 20596–20606, 2019.

[38] S. Zhou, J. Wang, D. Meng, X. Xin, Y. Li, Y. Gong, and N. Zheng, "Deep self-paced learning for person re-identification," *Pattern Recognit.*, vol. 76, pp. 739–751, Apr. 2018, doi: 10.1016/j.patcog.2017.10.005.

[39] F. Ma, X. Jing, Y. Yao, X. Zhu, and Z. Peng, "High-resolution and low-resolution video person re-identification: A benchmark," *IEEE Access*, vol. 7, pp. 63426–63436, 2019.

[40] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camera style adaptation for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5157–5166.

[41] X. You, J. Xu, W. Yuan, X.-Y. Jing, D. Tao, and T. Zhang, "Multi-view common component discriminant analysis for cross-view classification," *Pattern Recognit.*, vol. 92, pp. 37–51, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320319301074

[42] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 749–758.

[43] J. Wang, Z. Wang, C. Laing, C. Gao, and N. Sang, "Equidistance constrained metric learning for person re-identification," *Pattern Recognit.*, vol. 74, pp. 38–51, Feb. 2018.

[44] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: Problem overview and current approaches," *J. Ambient Intell. Humanized Comput.*, vol. 2, no. 2, pp. 127–151, 2011.

[45] V.-H. Diaz-Ramirez, J.-E. Hernández-Beltrán, and R. Juarez-Salazar, "Real-time haze removal in monocular images using locally adaptive processing," *J. Real-Time Image Process.*, vol. 16, no. 6, pp. 1–15, Dec. 2019.

[46] A. Ruchay, V. Kober, and J. A. Gonzalez-Fraga, "Reliable recognition of partially occluded objects with correlation filters," *Math. Problems Eng.*, vol. 2018, pp. 1–8, May 2018, Art. no. 8284123.

[47] D.-S. Bolme, B.-A. Draper, and J.-R. Beveridge, "Average of synthetic exact filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2105–2112.

[48] E. Santiago-Ramírez, J. Á. Gonález-Fraga, E. G.-L. O. Á. Xochihua, and S.-O. Infante-Prieto, "Face detection method based on non-linear composite correlation filters," *Res. Comput. Sci., Adv. Comput. Sci., Control Commun.*, vol. 69, pp. 215–226, Apr. 2014.

[49] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.

[50] E. Santiago-Ramírez, J.-A. González-Fraga, and S. Lázaro-Martínez, "Face recognition and tracking using unconstrained non-linear correlation filters," *Procedia Eng.*, vol. 35, pp. 192–201, May 2012, doi: 10.1016/j.proeng.2012.04.180.

[51] J. Valmadre, L. Bertinetto, J.-F. Henriques, A. Vedaldi, and P.-H.-S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2805–2813. [Online]. Available: http://arxiv.org/abs/1704.06036

[52] E. Santiago-Ramirez, J.-A. Gonzalez-Fraga, E. Gutierrez, and O. Alvarez-Xochihua, "Optimization-based methodology for training set selection to synthesize composite correlation filters for face recognition," *Signal Process., Image Commun.*, to be published.

[53] V. Diaz-Ramirez, A. Cuevas, V. Kober, L. Trujillo, and A. Awwal, "Pattern recognition with composite correlation filters designed with multi-objective combinatorial optimization," *Opt. Commun.*, vol. 338, pp. 77–89, Mar. 2014, doi: 10.1016/j.optcom.2014.10.038.

[54] B. Javidi, W. Wang, and G. Zhang, "Composite Fourier-plane nonlinear filter for distortion-invariant pattern recognition," *Soc. Photo-Opt. Instrum. Eng.*, vol. 36, no. 10, pp. 2690–2696, Oct. 1997.
[55] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Proc. CVPR*, Jun. 2011, pp. 81–88.
[56] M. Jones and P. Viola, "Fast multi-view face detection," Mitsubishi Electr. Res. Lab., Cambridge, MA, USA, Tech. Rep. TR2003-96, Aug. 2003. [Online]. Available: https://www.merl.com/publications/TR2003-96/
[57] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, 2nd ed., S. Z. Li and A. K. Jain, Eds. London, U.K.: Springer, 2011.
[58] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2009.

EVERARDO SANTIAGO-RAMIREZ received the bachelor's, M.S., and Ph.D. degrees in computer science from the Universidad Autónoma de Baja California, Ensenada, Baja California, México. He is currently a Professor and a Researcher with the Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, México. His research interests include person reidentification, face recognition, computer vision, and correlation filters.

JOSE ANGEL GONZALEZ-FRAGA received the B.Sc. degree in electrical engineering from the Universidad Autónoma de San Luis Potosí (UASLP), México, in 2002, and the M.Sc. and Ph.D. degrees in computer science from the Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), México, in 2004 and 2007, respectively. He is currently a full-time Professor with the Universidad Autónoma de Baja California. His research interests include pattern recognition, adaptive image processing, and robot vision.

EVERARDO GUTIERREZ LOPEZ received the Ph.D. degree in computer sciences from the Centro de Investigación Científica y Estudios Superiores de Ensenada (CICESE), México, in 2010. He is currently a Faculty Member and a Researcher with the Facultad de Ciencias, Universidad Autónoma de Baja California, Ensenada, Baja California, México. His research interests include combinatorial optimization, heuristic algorithms, and bioinformatics.

OMAR ALVAREZ-XOCHIHUA received the Ph.D. degree in computer science from Texas A&M University, USA. He is currently a Professor of computer science with the Universidad Autónoma de Baja California, México. He is conducting research activities in the areas of educational technology, knowledge representation, and natural language processing.

JUAN ACOSTA-GUADARRAMA received the Ph.D. degree from TU-Clausthal. He is currently a Full Research Professor with the Autonomous University of Juarez. His research interests include mathematical logic, theoretical computer science, and knowledge representation and reasoning. He is a member of the Mexican Association for Artificial Intelligence (SMIA) and has been appointed as National Scientist by CONACyT.