Applied Soft Computing Journal 89 (2020) 106014

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Energy-efficient face detection and recognition scheme for wireless visual sensor networks

Abdulaziz Zam, Mohammad Reza Khayyambashi *, Ali Bohlooli *

Faculty of Computer Engineering, University of Isfahan, Hezar Jerib Ave., Isfahan 81746-73441, Iran

ARTICLE INFO

Article history:
Received 6 April 2019
Received in revised form 24 November 2019
Accepted 9 December 2019
Available online 15 January 2020

Keywords:
Visual sensor network
Detection-accuracy and energy-consumption trade-off
Energy efficient face detection
Distinctive vector
Face recognition

ABSTRACT

Energy-efficient and robust face detection and recognition scheme can be useful for many application fields such as security and surveillance in multimedia and visual sensor network (VSN). VSN consists of wireless resources-constrained nodes that are equipped with low-energy CMOS cameras for monitoring. On the one hand, captured images are meaningful multimedia-data that impose high energy consumption to be processed and transmitted. On the other hand, visual sensor (VS) is a battery-powered node with limited life-time. This situation leads to a trade-off between detection-accuracy and power-consumption. This trade-off is considered as the most major challenge for applications using multimedia data in wireless environments such as VSN. For optimizing this trade-off, a novel face detection and recognition scheme has been proposed in this paper based on VSN. In this scheme, detection phase is performed at VS and recognition phase is accomplished at the base station (sink). The contributions of this paper are in three folds: 1. Fast and energy-aware face-detection algorithm is proposed based on omitting non-human blobs and feature-based face detection in the considered human-blobs. 2. A novel energy-aware and secure algorithm for extracting light-weight discriminative vector of detected face-sequence to be sent to sink with low transmission-cost and high security level. 3. An efficient face recognition algorithm has been performed on the received vectors at the sink. The performance of our proposed scheme has been evaluated in terms of energy-consumption, detection and recognition accuracy. Experimental results, performed on standard datasets (FERET, Yale and CDnet) and on personal datasets, demonstrate the superiority of our scheme over the recent state-of-the-art methods.

1. Introduction

VSN, the new generation of wireless sensor network (WSN), has been developed recently due to the fast evolution of tiny-scaling technology and embedded systems. Each visual sensor (VS) is equipped with a camera and has capabilities of capturing, saving, processing and transmitting frames [1]. In addition to having acoustic motion-sensor, some or all VSN-members have visual capturing and processing capabilities. These capabilities encouraged researchers to transfer visual applications from wired networks to VSN to facilitate surveillance in difficult access areas [2,3]. There are, in addition to energy constraints, memory, processing and bandwidth limitations in VSN resulting in complex challenges to be tackled for efficient handling of multimedia-data. Face detection and recognition is considered as an essential application for in-door and out-door surveillance systems. Face detection process itself is a complex challenge for the vision-based automated systems because individual class (face) should be isolated from a large number of classes presented in each captured frame [3,4]. Most of face detection algorithms have been developed essentially for operating in a centralized system or in wired camera networks where there are rich-resources of power, bandwidth, memory and processing abilities [5,6]. Since traditional algorithms tend to obtain maximum detection-accuracy in wired environments and do not consider power conservation mechanisms, employing these algorithms directly in resources-constrained VSN is an inefficient approach. In this paper, an energy-aware face-detection algorithm is proposed at VS. The proposed algorithm uses a novel approach for detecting faces and extracting distinctive-vectors from the detected faces based on histogram of oriented gradient (HOG) features [7].

Since the camera of each VS is static, it maintains relatively stationary background (little changes in illumination and shadows related to sun-light situation in outdoor) of the monitored scene. Background would be changed by an entry or movement of any object into the scene. In our proposed scheme, face-detection phase is carried out in resources-constrained VS and hence, searching for faces is restricted only to the changed portion of the frame. According to this restriction, sliding search-window would scan a few number of boxes (instead of searching

* Corresponding authors.
E-mail addresses: a.zam@eng.ui.ac.ir (A. Zam), Khayyambashi@comp.ui.ac.ir (M.R. Khayyambashi), Bohlooli@eng.ui.ac.ir (A. Bohlooli).

https://doi.org/10.1016/j.asoc.2019.106014
1568-4946/

the whole frame), which leads to reducing the detection-delay and in-node energy-consumption. When an object enters the scene causing background changing, corresponding-VS runs background subtraction algorithm to extract foreground-blobs by subtracting corresponding pixels of two sequential frames considering a threshold value. In this paper, adaptive Gaussian mixture probability density model [8] is performed due to a refined approach for modeling the background and extracting blobs. The proposed approach maximizes the number of Gaussian mixture components (GMC) in the face-block of each extracted blob to achieve high accuracy of face-detection in this part (block) of the target image.

For further energy-efficiency improvement, rather than searching all extracted blobs, our algorithm omits non-human blobs that have negligible probability of containing a face. Since humans have convergent values of their dimensions-ratio, we use a simple width/height-ratio feature of the extracted blobs to omit non-human blobs. Accordingly, search-window would be focused only on the most important part of the frame, which has maximum probability of containing human. Furthermore, search-window starts scanning from the face-expected block of the considered-blob in which, GMC is maximized for efficient detection. For face detection in a considered blob, we use a supervised learning method by extracting histogram-of-oriented-gradient features of sliding search-windows and comparing them with stored training features. Support vector machine classifier (SVM-one-class) is used to separate the face-class from non-face-classes by a desired kernel function. After detecting faces, VS should send detected-faces to sink for further processing and recognition.

In VSN, high percentage of the residual energy is consumed due to the communications between nodes (transmitting and receiving). Accordingly, for conserving energy and reducing recognition-delay, instead of sending raw data of the detected-faces, discriminative-vectors would be abstracted from already extracted-features of the detected-faces to be sent to sink. This novel approach contributes to improving the performance in three-dimensions: 1. Discriminative vectors have lower size compared to raw data and hence, occupy lower bandwidth and consume less energy for transmitting and receiving. 2. Discriminatory vectors are invariant and more distinctive data that facilitate further processing (e.g., recognition) when being received at sink. 3. Extracted vectors are encrypted data to be communicated that increase the security level of the VSN.

Sink is connected to a renewable power supply and has super capabilities of processing and high capacity of memory. Accordingly, discriminative vectors with SVM classifier have been used in a comprehensive manner for recognition of the received faces. For face recognition, sufficient and multi-scaled collection of datasets have been used in training for achieving maximum possible recognition accuracy [9].

The rest of this paper is organized as follows; in Section 2, the most related studies are presented and discussed. Section 3 illustrates the proposed scheme in details. Simulations and evaluations of the proposed scheme are presented in Section 4 and finally, in Section 5, the conclusion of this study and suggestions for future works are presented.

2. Related works

Face detection and recognition in VSN is a more challenging procedure than which is considered in wired systems. Since detection-accuracy | power-consumption trade-off should be considered in this resources-constrained environment, conventional algorithms do not work efficiently for optimizing this trade-off. However, this topic is recently gaining interest and there are only few approaches have been suggested for object or face detection in VSN striving to reduce in-node and in-network energy consumption by maintaining acceptable accuracy. In [10], an energy-efficient approach has been proposed for object detection and image transmission in wireless multimedia sensor network (WMSN: is a special type of VSN). In this approach, each camera node (CN) captures its initial background, stores and sends it to sink. CN should divide each captured image to 4-blocks. When a new object or new movement appears in a captured image, CN determines the particular area (block) in which changes occurred and sends only changed block to the sink in which the whole frame will be reconstructed by the reference frame received earlier. Although this approach has achieved noticeable in-node and in-network energy conservation, it still needs to transmit block-images in their raw pixel-domain. Moreover, its performance would be degraded in the case of multi-objects appearing simultaneously in spaced areas of the frame, the case which demands sending most of the frame-blocks.

In [11], a dynamic approach has been suggested for enhancing the performance of face-detection in VSN. In this approach, background subtraction and image reduction algorithms have been used for bounding detected objects in box-shapes, and then dynamic approach would be performed for detecting and cutting faces from these boxes. Finally, faces would be sent, in their pixel level, to the sink for any further processing. This approach uses probabilistic computations to predict the location of the face in the human bounding-boxes. However, predicting the location of the face causes non-negligible errors (including false-positive and true-negative samples). Moreover, sending pixels of detected faces would affect the efficiency of power consumption.

In [3], an energy-aware method has been proposed for detecting faces and prolonging the life-time of the visual sensors. In this method, captured images are preprocessed at VS to be prepared for face-detection algorithm, which extracts Haar-like features and uses boosting-based classifier to detect faces. Detected faces will be sent to sink for recognition. This approach of classification combines multi-classifiers in a cascade structure and omits non-face boxes at each stage of the cascade. Although this classification method is faster, it has lower accuracy than SVM classification. Although this method has achieved relatively high in-node energy saving due to its fast and energy aware local processing and classification algorithms, it still has a shortage of sending raw data of the bounding-boxes, which imposes high in-network energy consumption and bandwidth conservation. In the paper, authors have evaluated their work in term of energy consumption using 120×120 m of network-size and 25-VS nodes only, whereas the effect of multicasting raw-data of images through the network would appear clearly by extending the network size and increasing the number of VSs as will be described in Section 4.2.2 of this paper. Moreover, the detection accuracy of using fast classification method has not been discussed in the paper.

Since detected faces should be sent to sink for recognition, which requires features-extraction of received data, there is no reason to send raw data to sink ignoring the already extracted features of the same data. In the case of having another reason for sending raw data (e.g., target or up-normal behavior detection), then it can be done on-demand, upon the request of the sink to a specific VS as proposed in our previous work [12].

In [13], authors have proposed a system architecture based on wireless multimedia sensor network for classification of moving objects at VS and sending the results to sink. This architecture has been implemented using designed hardware components on recorded-videos dataset. They have used simple shape-based features with SVM classifier for classification of detected objects and sleep/active modes are considered for energy conservation. In this

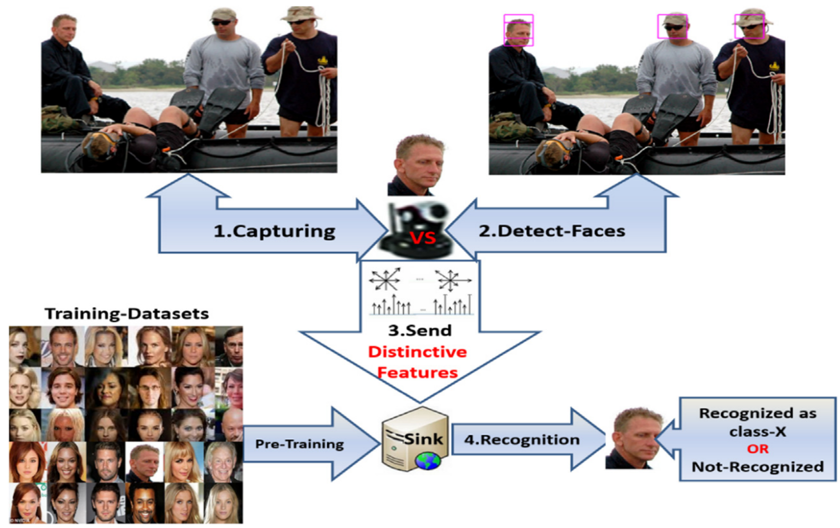

Fig. 1. The overall system operation.

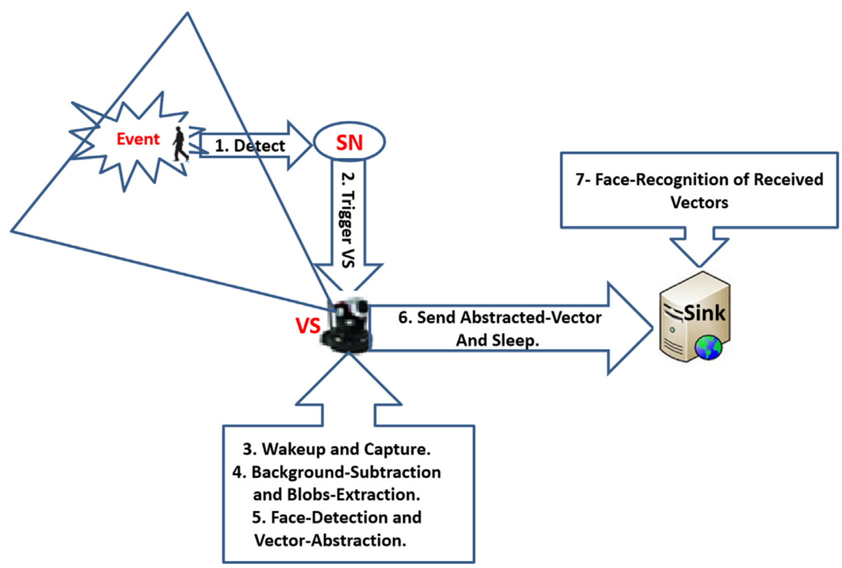

Fig. 2. Block diagram of the proposed scheme procedures.

work, there is no further processing (e.g., recognition) considered at sink and all procedures are accomplished at VS, which imposes heavy burden on these resources-constrained components shortening their life-time.

Using supervised learning method for face-detection depending on invariant and distinctive features still achieves better performance. Furthermore, abstracting discriminatory vectors of detected-faces to be sent to sink, the approach that has not been studied in the previous works, makes a big step towards security, robustness and energy-efficiency as will be proven in Section 4.2 of this paper. We have evaluated our proposed scheme and compared our results with all aforementioned methods demonstrating the superiority of the proposed scheme above these state-of-the-art methods.

3. Proposed scheme for face detection and recognition

As shown in Fig. 1, the overall operation of our proposed scheme is presented. Initially, all VSs would set the sleep-mode conserving their residual energy. SNs with very low energy consumption have the responsibility of movement detection in the network area. When any SN detects a movement, it should trigger corresponding VS to wake up and start visual detection. Triggered VS would return to sleep-mode after completing the vector-extraction procedure as described in algorithm.1. As shown in Fig. 2, the operation starts at triggered VS by: capturing images from its field of view (FoV) and comparing them with the last updated background. If there is any considerable change, adaptive background subtraction algorithm will be run to extract appearing blobs. Faces would be detected from the considered-blobs using SVM classifier and HOG-features. The already extracted features of each face should be abstracted (using our novel approach) in one discriminative vector to be sent to sink for recognition. After sending abstracted-vector, VS should set the sleep-mode. At sink, voluminous training features are already extracted from large datasets and stored in the local memory due to offline supervised training procedure. Finally, by receiving discriminative-vector of any face at sink, this face would be classified and recognized using SVM classifier. These steps of our scheme will be detailed and explained separately in the next subsections. In this proposed scheme, energy-aware approach has been proposed, which is accomplished through two-phases:

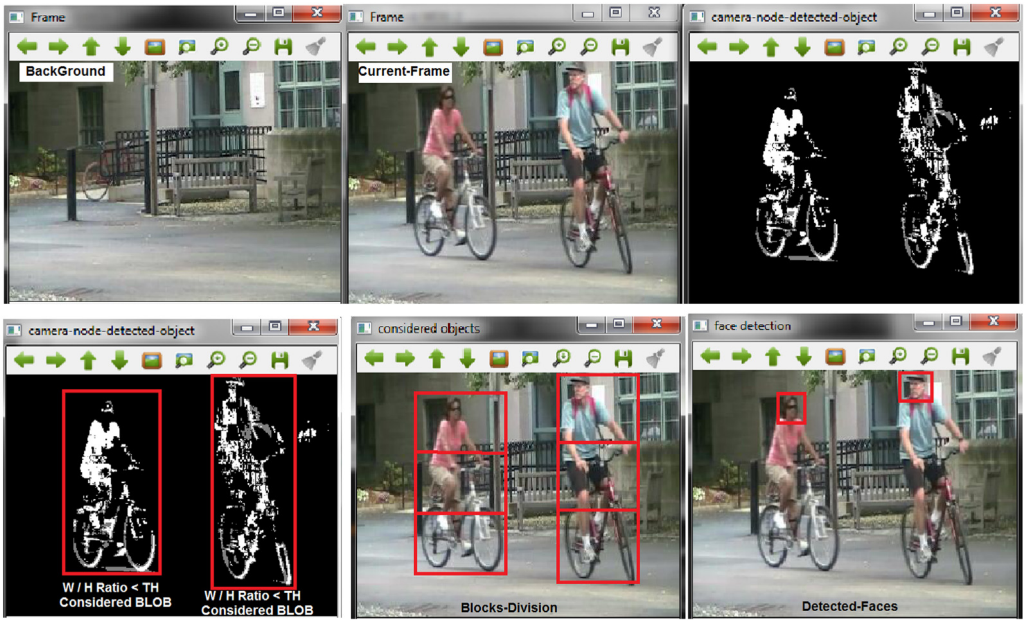

Fig. 3. Blob-consideration, blocks-division and face-detection (backdoor-dataset).

detection phase, which is conducted due to energy-aware manner because this phase is accomplished in energy-constrained VS nodes. And recognition phase, which is accomplished in comprehensive and performance-starving manner for attaining the best possible accuracy as this phase is accomplished in rich-resources base station (sink).

3.1. Energy-aware face detection at VS

Our proposed face-detection algorithm tends to efficiency by avoiding transmission of spatio-temporal redundant data and by abstracting light-weight but distinctive information from selected multimedia data. Initially, each VS captures an initial-background image from its FoV, stores and sends a cloned image of the initial-background to sink. Then all VSs set the sleep-mode to conserve their residual energy.

When any movement occurs in any detection-range, changes should be acoustically detected using acoustic motion-sensors (SNs). Accordingly, corresponding VS would be triggered by its own SN or by neighboring SNs as presented in [13]. Triggered-VS wakes-up and sets active-mode to start visual-detection procedure to visually detect appearing objects as shown in Fig. 2. Visual-detection initiates by the background-subtraction algorithm for extracting new appearing blobs in the scene. In this paper, adaptive Gaussian mixture probability density model (AGMM) is employed for background subtraction. AGMM imposes low computational cost and high resistance against illumination changes achieving good accuracy. In AGMM, the number of Gaussian mixture components (GMC) can be set adaptively according to the pixel-changes frequency. Supposing that $(X_{i,t})$ is the current value of pixel_i at time_t and $(u_{i,t-1})$ is the previous average value for the same pixel_i, the following equation is used for determining changed pixels (foreground). Pixel_i is considered as foreground-pixel according to the formula (1). Otherwise, it is determined as background. TR is a threshold value and it is set to 10 in our algorithm.

$$\left|X_{i,t} - u_{i,t-1}\right| > TR \tag{1}$$

Background should be updated continuously that a new appearing object may become a part of the updated background. Background is updated according to the following formula (2):

$$B_i = (1 - r) \times I_i + r \times B_{i-1} \tag{2}$$

B_{i-1} is previous background, B_i is current updated background, I_i is current captured image and r is updating rate. The output of foreground-extraction algorithm is a binary (black | white) image in which, extracted blobs are {1s} (e.g., white) and the background is {0s} (e.g., black). It is easy to detect edges in binary images and thus, width/height ratio of each blob can be easily calculated. Based on W/H-ratio, the algorithm omits non-human blobs that have W/H-ratio greater than threshold. In this algorithm blobs-threshold is set to (TH = 0.5). Obviously, non-human objects such as cars, bicycles, animals or buses have W/H-ratio greater than 0.5 and accordingly, these objects would be omitted. Whereas human blobs would be considered even if these humans are riding bicycles as shown in Fig. 3. Human-blobs consideration procedure is described in Algorithm.1.

The next step is detection of the faces presented in considered blobs. As described in algorithm.2, each considered-blob of human is divided horizontally to 3-equal-blocks. This blocks-division spatially facilitates the face-detection process due to two-folds: 1. Search-window should scan the first top-block and if a face is detected, scanning procedure should stop to conserve residual-energy of VS. 2. The number of GMC-components would be maximized in the top-block, which has high probability of containing a face. This GMC-maximization makes the Gaussian-model able to detect small changes occurring in this area and classify changed pixels as foreground. Since GMC-maximization imposes high computational-cost, the proposed algorithm gets the benefit of GMC-maximization efficiently by focusing on a small portion (block) of the image. The algorithm can detect the same face through consequent-frames by capturing the face in different positions. For ensuring that the detected-faces in the sequence belong to the same individual, corresponding face-bounding-box would be tracked for a short time using median-flow tracker. Accordingly, various sets of feature-vectors will be extracted from the same face, which improves the accuracy of recognition at sink. For detecting the face presented in the top-block of each considered-blob, features are extracted from the top-block and swept with pre-trained SVM classifier to localize

***Algorithm.1*:** main block of the application at VS

***Input*:** trigger signal from motion sensor SN
***Output*:** human blobs $\boldsymbol{B_i}$ being saved locally and face sequences $\boldsymbol{F_i}$ being sent to sink

If (Triggered == "Yes") // *start on motion-detection*
  Wakeup()
  Capture()
  Update_background( ) // *based on formula.2*
  Foreground_extraction()
  Extract_Blobs()
  For each Blob // *from i = 1 to i = extracted_blobs_number*
    Ratio = Blob.Width / Blob.Height
    if (Ratio ≤ Threshold) // *(Threshold = 0.5)*
      Blob.save() // *save the blob in VS local memory*
      $\boldsymbol{B_i}$ = Blob
      Check_times = 0 // *checking for a face will begin at the next line*
      Get_sequence($\boldsymbol{B_i}$) // *algorithm.2: gets $\boldsymbol{B_i}$ and returns $\boldsymbol{F_i}$*
      Vector_extract($\boldsymbol{F_i}$) // *extract $\overrightarrow{\boldsymbol{V_i}}$ from images-sequence $\boldsymbol{F_i}$*
      Vector_send($\overrightarrow{\boldsymbol{V_i}}$) // *send extracted vector ($\overrightarrow{\boldsymbol{V_i}}$) to sink*
    Else Blob_omit() // *omit non-human blobs*
    End if
  End For
  Sleep() // *return to sleep mode after completing visual detection*
Else Sleep() // *continue sleeping if not triggered*
End If

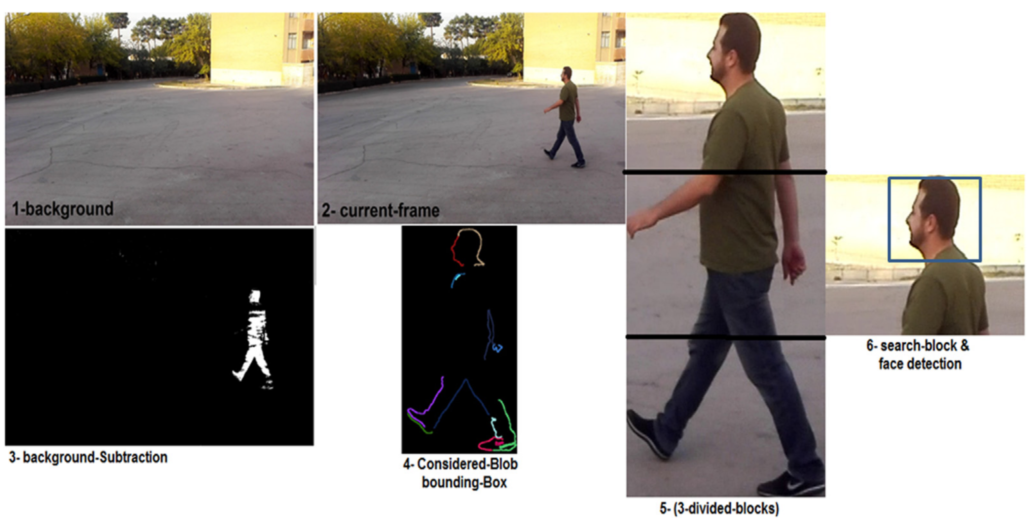

Fig. 4. Blob-consideration, blocks-division and face-detection (personal-recorded-video).

the face in the block-image. SVM classifier is suitable for fast separation of two-classes (in this case: face-class and non-face-class) and can perform well even with one positive-sample. More details of the methodology of this classifier will be described in the next sections.

3.2. Discriminative-vector extraction from detected face sequence

At this step, discriminative vector of detected-face sequence should be extracted at VS and sent to sink for further processing and recognition. Discriminative vector of our proposed scheme would be abstracted based on the already extracted features from face-sequence using a novel approach. Features are extracted based on the histogram of orientation (HOG). HOG is still one of the best choices for face detection and recognition because of its robustness against variations (even in faces with large variations in appearance and illumination changes) [14]. The main insight of the HOG approach is that the local object shape within an image can be described by its edge directions. HOG descriptor is specified by two values: the magnitude of the gradient and the orientation of the gradient. As shown in Fig. 5, the image is divided to connected cells. Each cell contains number of pixels (e.g., 4×4 pixels in each cell). For each cell, normalized histogram of orientations is compiled using the aforementioned two-values of each pixel within the cell. For improving robustness against illumination, shadowing and contrast changes, cells are grouped in larger spatially overlapped regions (blocks). Then, gradient magnitudes are normalized at each block.

Algorithm.2: face detection, tracking and get sequence

Input: human blob B_i
Output: face sequence F_i(number of face-images belong to the same individual)

```
Get_sequence( B_i)
{
Check_times = Check_times + 1
If (Check_times == 1)
    Blob_divide() // divide the blob B_i to 3-Blocks only for the first check time
End If
Block_features_extract()//extract HOG features from top-block where face is expected
Label = SVM_Predict(HOG) //classify the object in the block (face or non-face)
If (Label == "Face")
   Tracker.setup()  // setup the tracker for the face-block
 Else  If (Check_times < 3)
          Wait()
          Next_frame.read()
          GMC = Max // components maximization for efficient changes-detection
          Block_update_background( ) // update the background based on Max GMC
          Get_sequence( B_i) // efficiently repeat checking the block
      Else Return (No-face) // exit when no face detection after 3 attempts
      End If
End If
While (Block.tracked == "yes" and Sequence.length < 5) // 4-images per sequence
     Next_frame.read()
     Next_frame.track()
     New_box = Tracked_box.cut()  // cut the face-image to be added to the sequence
     F_i =  F_i + New_box  // adding the image to the sequence
     Wait() // delay is related to frame capturing rate of the camera
End while
Return( F_i )
} // end function Get.sequence( B_i)
```

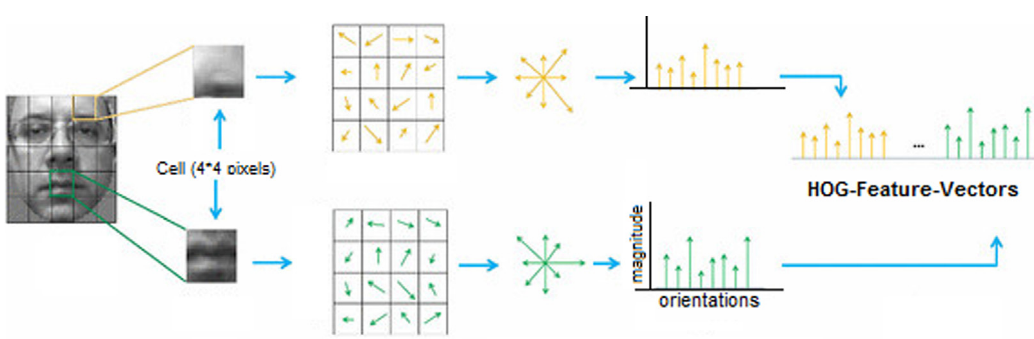

Fig. 5. HOG-features extraction and vectors concatenating.

In the previous section, we have mentioned that the considered blobs are divided to 3-blocks and the number of GMC-components is maximized at the first top-block where the face is expected. This GMC-maximization helps AGMM to provide effective modeling for this part of the background. Accordingly, any changed pixel in this block, due to a movement of the face, will be classified as foreground in the next frame. Considered face-bounding-box would be tracked by median flow tracker, which estimates the motion of the box between consequent frames by predicting the displacements of a number of points within the bounding box. This tracker guarantees that detected-faces within the tracked box are belonging to the same individual. And hence, the algorithm will be provided with images-sequence representing the same face in many positions. From this sequence, a number of images (e.g., 4) will be selected to produce unique discriminatory vector. As shown in Fig. 6, features should be extracted from the selected face-samples in the sequence after preprocessing (including gray-scale conversion and equalization). And then, for each sample, features-vector is concatenated. These extracted vectors contain comprehensive information of the face that are important for accurate recognition. For instance, this information may include multi-scale features (when the face gets closer to the camera between capturing-intervals), the case that can compensate for the weakness of HOG against scale-variations. However, since these frames are captured due to short intervals,

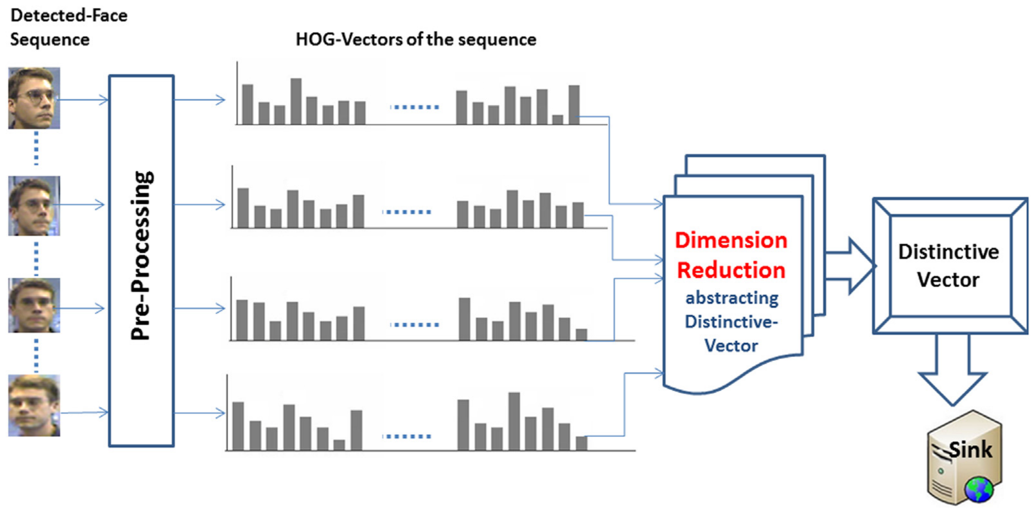

Fig. 6. Extraction of feature-vectors from face-sequence and abstraction of distinctive vector.

there are spatio-temporal redundant pixels among the frames of the sequence. Accordingly, extracted histograms should have frequently repeated values.

The target of our novel proposed approach is abstracting light-weight but discriminative information, which can represent the uniqueness of the face. On the one hand, the uniqueness of this information will improve the performance of recognition at sink. On the other hand, light-weight information imposes low cost of transmission conserving in-network energy. In the vision-based systems, the information, which has lower frequency is more discriminative and much important. Whereas information, which has high frequency is considered as unimportant background (imagine an image represents a small bouquet of various flowers in front of a large unified wall). Therefore, extracted vectors of the sequence should be dimensionally reduced due to the followings; firstly, redundant irrelevant vectors (vectors that have the same voting value (y) at the same orientation (x)) would be removed from histograms, which aids to provide suitable input-space to the principal component analysis (PCA) in the second step. Secondly, after omitting redundant vectors, the popular dimensions reduction model PCA [15] is applied to the remaining vectors to minimize the correlation between vectors and abstract one discriminative vector. PCA is the most compatible dimension reduction method for face-recognition applications [16]. PCA keeps the most discriminative features (eigenvectors) by maximizing the covariance of the data in the output (low-dimension) achieving effective discrimination for classification. In our approach, rather than applying PCA only on intra-vector of each face-image, we have also applied it on inter-vectors between the face-images of the sequence. In our approach, the covariance matrix and then eigenvectors and eigenvalues are computed firstly for each vector per image as described in formulas (3) and (4).

The eigenvectors with highest eigenvalues are selected retaining 20-percent of the features per image namely image-vector. And then, the covariance of selected image-vectors per sequence is also computed and finally retaining 10-percent of the data namely sequence-vector. Since the redundancies in inter-vectors are much more than the redundancy in intra-vector, we have selected lower percent of the features for concatenating the distinctive-vector of the sequence.

$$Cov\,(\text{Magnitude},\ \text{Orientation}) = \frac{\sum_{i=1}^{n}(M_i - \overline{M}) \times (O_i - \overline{O})}{(n-1)} \tag{3}$$

$\overline{M}$ is the average vector of the Magnitudes and $\overline{O}$ is the average vector of the orientations:

$$\overline{O} = \frac{1}{N}\sum_{i=1}^{N} O_i \text{ and } \overline{M} = \frac{1}{N}\sum_{i=1}^{N} M_i \tag{4}$$

In wireless sensor networks, communications (transmitting and receiving data) impose much higher cost, including high energy-consumption, than in-node local data processing [12,17]. Therefore, in this proposed scheme, we have somewhat increased local processing of multimedia data to abstract light-weight but discriminative information for the sake of reducing communications-cost and increasing the security-level of the network. As shown in Fig. 6, the output (discriminative vector) of the local-processing should be sent from VS to sink. In this paper, we have used X-Bee communication protocol for transmission on IEEE 802.15.4, non-beacon CSMA-mode radio model. This is the energy-efficient choice of the existing communication protocol-models for VSNs.

3.3. Face recognition of the received distinctive-vectors at sink

Once a discriminative vector is received at the sink, its corresponding face should be recognized. Initially, large and appropriate collections of training samples should be used for training. Training datasets should include labeled samples that are suitable for the target area of interest. For instance, if VSN is covering a military area, training datasets should contain samples of all soldiers that are expected to wander in the coverage area. In this paper, we have utilized standard datasets (FERET database in [18], Yale database in [19] and CDnet dataset in [20]) for evaluating our proposed scheme. Block diagram of the recognition algorithm at sink is shown in Fig. 7. Recognition algorithm uses HOG-based discriminative-vectors extraction method for extracting training features from the training samples of the utilized dataset. Since training procedure is accomplished "offline" and there is no limitations bounding power, memory and processing costs at sink, multi-scale features extraction of training samples helps for achieving better recognition-performance. In [21], comprehensive comparison study of the existing features-extraction methods has been presented based on comprehensive experimental sessions. In this study, it has been proved that HOG-descriptor is one of the most suitable descriptors for characterizing facial expression peculiarities. In this paper, we have boosted HOG-descriptor with multi-scale features and PCA dimension reducer to get the most

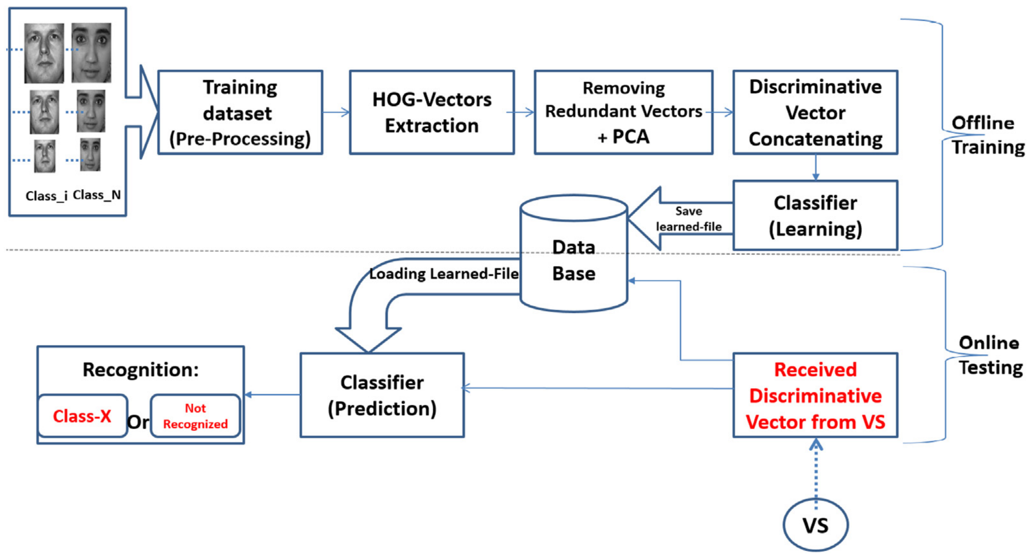

Fig. 7. Block diagram of the proposed recognition algorithm at sink.

salient points of the face to improve the performance of the classification.

For obtaining best results due to measuring the vector-distance between test and training vectors, using multi-scale samples for training at the sink should be compatible with the discriminative-vector extraction method used at VS. For this reason, we have used down-sampled face-images for training. Because the information content of the up-sampled image is almost the same as that is in original image, whereas down-sampling causes loss of high frequency information [22]. Accordingly, the important (low-frequent) information will be remained in both test and training discriminative-vectors causing matching measurement of the vectors-distances.

As shown in Fig. 7, labeled vectors are used by the classifier through "offline" learning procedure, which aids the classifier to categorize a new unlabeled sample in the right class through "online" testing procedure. In this algorithm, we have utilized SVM classifier, which formalizes a hyper-plane to separate a set of positive samples from a set of negative samples with maximum margin in the training step. Accordingly, SVM is appropriate for two-class separation problems. Since we have multi-class of faces, SVM has been performed for each pairwise of classes separately "one-against-one" as introduced in [23].

From many kernel functions that are usually being utilized for SVM classifier, we have utilized radial basis function (RBF). RBF provides non-linear separation of the data. It computes the squared Euclidean distance between two feature-vectors and takes the negative exponential of this distance to quantify the similarity measure as a value between zero and one. From existing kernels in SVM, RBF mostly obtains the highest accuracy. Given two feature-vectors $x1$ and $x2$, RBF similarity value is computed due to the following formula where σ is standard deviation;

$$RBF\ (x1, x2) = \exp(-\frac{||x1 - x2||^2}{2\sigma^2}) \tag{5}$$

For predicting the right class for a new received test-sample at sink, voting system is used to select the appropriate class. A number of learned classifiers (with different parameters) contribute in predicting the right class for a test sample. Each classifier gives a vote for one of two-classes and the final decision is adopted based on the majority of the votes.

4. Experimental results and performance evaluation

For evaluating the performance of the proposed scheme, we have employed OpenCv library with QT programming environment and C++ language for analyzing the performance in terms of detection accuracy and correct recognition rate. Besides, OMNET++/Castalia provided in [24], the visual sensor network simulator, has been used for evaluating in terms of injected traffic, in-node and in-network energy consumption. OMNET++/Castalia is a new extension framework, which provides suitable tools for simulating VSN due to performing image encoding and transmission of real image packets. Besides, this framework performs real energy measurements from low-cost image sensor built around an Arduino board for realistic energy parameters. All experiments are performed on FERET, Yale and CDnet databases in addition to personal recorded videos. CDnet is an exhaustive change-detection benchmark database in which, videos are categorized based on the target application and the level of challenges. As our scheme is dealing with humans-detection, we have selected pedestrians-detection datasets from the categories for evaluating our scheme. From pedestrians-datasets we have selected two high-level challenging datasets: 1 — backdoor-dataset: as can be shown in Fig. 3, it has fast moving objects, soft shadows and soft background changing. 2 — bus-station-dataset: as shown in Fig. 8, it has multi-objects moving, hard shadows and dynamic background. FERET and Yale datasets are considered for evaluating the recognition accuracy. Yale-dataset contains 15-classes with 11-face-images for each class taken under different lightning conditions and different facial expressions. Yale is hard challenging face-recognition dataset compared to FERET in which, the images are taken under semi-controlled conditions with no lightning changes. Our personal recorded videos are divided into two parts, low quality dataset of 320×240 pixels and high quality dataset of 1920×1088 pixels, which both contain cars and pedestrians moving, soft shadows and hard background changing due to tree leaves vibration and half shaded area as shown in Fig. 4.

4.1. Parametric optimization

The values of the internal parameters of features-extraction and classification algorithms impose direct effects on the performance. Accordingly, careful selection approach of these values should be used based on the experimental results. HOG descriptor

Fig. 8. 6-Separated-frames of the sequence of bus-station dataset.

is characterized by three parameters; cell size, block stride and the number of bins. #Bins: is the number of the orientations partitioning the angular range (from 0° to 360° or from 0° to 180°) to quantify the orientation of each vector in a vectors-histogram. Smaller cell-division approach leads to higher number of vectors in the histogram representing more details of the image and vice versa. Likewise, high number of bins leads to high oriental representation and crowded histograms and vice versa. The block stride refers to the number of times in which a cell should be normalized for obtaining resistance against some variations. In this proposed scheme, since the dimensions of the output features would be reduced later and based on many experimental sessions, HOG parameters that have achieved the best results are as followings; the number of bins is set to 8, cell size is set to 8×8 pixels, block size is 2×2 cells and block stride is over 8 pixels for face detection. For face recognition at sink, the number of bins is still 8, cell size is 6×6, block size is 2×2 cells of 6×6 pixels and block stride is over 6-pixels.

The influential parameter of dimension-reduction algorithm is explained-features covariance whose value ranges from zero to one. Since PCA sorts the vectors based on their covariance descending from the largest covariance vector to the smallest one, the value of this parameter refers to the percentage of the larger covariance vectors that remain in output. Using tough approach (e.g., remaining 75% of the vectors) leads to high correct matching rate but, in return, increases the computational cost and vice versa. In this proposed scheme, tough approach is performed for recognition at sink obtaining the best possible performance. Whereas, conformist approach (remaining 55% of the vectors) has been performed for face-detection at VS to avoid energy depletion. Accordingly, explained features covariance is set to 0.55 at VS and 0.75 at sink.

Finally, the parameters of the SVM-classifier should be selected. The kernel of SVM utilized in our scheme is RBF. RBF has two parameters known as penalty factors (C and gamma). The value of gamma ($\gamma = -\frac{1}{2\sigma^2}$) defines how far the influence of a single training sample reaches and C determines the boundary of the class. By other words, gamma is for centroid samples and C is for other surrounding samples. For face recognition at sink, large value of C is considered to aid training classifier to classify all training samples correctly. And intermediate value of gamma is considered to give more freedom to the classifier and at the same time to fit with a large value of C; (C=1024, $\gamma = 0.07$). For face detection at VS, since there are less amount of training samples, smaller value of gamma is considered. For face detection, only one learned classifier with RBF kernel ($\gamma = 0.03$, $C = 1573$) is performed. Whereas for recognition at sink, a number of SVM classifiers with deferent kernel parameters of (C: from 500 to 1500) and (γ: from 0.01 to 0.1) should be trained and contribute in a voting system to predict the right class for a new sample.

4.2. Experimental results

In this subsection, the performance results of our proposed scheme are evaluated, analyzed and compared to other recent related works. The evaluation is performed through three phases: 1. Evaluation of human and face detection algorithm at VS. This evaluation is performed in terms of precision and recall criteria. The precision of the algorithm is (P = TP / (TP + FP)). TP: true positive where the case is positive (face) and predicted as positive. FP: false positive where the case is negative (non-face) but predicted as positive (face). The recall of the algorithm is (R = TP / (TP + FN)) where FN: false negative where the case is positive (face) but predicted as negative (non-face). 2. Evaluation of power-consumption and injected-traffic efficiency of the face-detection algorithm in the network. 3. Evaluation of the performance of the recognition algorithm at sink in term of correct recognition rate, CRR = the number of test-samples (faces) that are classified correctly over the number of all test-samples. All our experiments have been performed on a computer of (windows-7, 4GB of RAM, Intel(R) dual core of 2.2 GHz).

4.2.1. Experimental results of the human and face detection algorithms

From pedestrians-detection database in [20], we have chosen backdoor-dataset, which contains 2000 frames with dimensions of 320×240 and bus-station dataset, which contains 1250 frames of 360×240. These selected data-sets mostly contain pedestrians with challenging conditions of shadowing and fast objects-moving. Initially, the first step of our detection algorithm has been performed for human-detection based on foreground subtraction and W/H-Ratio calculation without any training or classification. The experiments have been repeated 15-times using 100 test-images from backdoor dataset and 100 test-images from bus-station dataset. We have compared our results of humans-detection accuracy with the results obtained by [10] in which, the same datasets have been utilized where the moving objects are mostly pedestrians. As presented in Table 1, our approach has achieved better accuracy of humans-detection.

Table 1
Comparison results of human-detection accuracy.

Approach	Dataset	Recall	Precision
Our scheme	Pedestrians	61%	68%
Approach of [10]	(backdoor)	56%	64%
Our scheme	Pedestrians	65%	74%
Approach of [10]	(bus-Station)	58%	65%
Our scheme	Personal data	63%	71%
Approach of [13]	(320×240 pixels)	48%	57%

For fair evaluation and comparison, we have compared the results of our algorithm on recorded videos (320×240 pixels) with the results of approach [13], which uses recorded videos with the same dimensions and we have also achieved better accuracy as presented at the bottom of Table 1.

In [10], blocks containing moving objects are detected at VS and sent to sink at which, the whole frame will be reconstructed and new appearing objects can be classified. Disregarding erroneous wireless links, this method imposes delays of delivering detected blocks to sink before classification. It also causes reduction of the images-quality due to transferring from VS to sink, which affects classification-performance at the sink. In our approach, humans are detected at VS using simple calculations with very low computational-cost.

Then, using opencv-3.2 library with the parameters defined in Section 4.1 for face-detection, 30-simulation sessions have been performed on the human-blobs that are detected from the previous step. For learning the classifier, from each dataset (bus-station and backdoor) 400-cropped-images (do not contain faces) are selected as negative training samples. 150-cropped-face images collected from face detection datasets are utilized as positive training samples. Besides, recorded video with 1920×1088 resolution has been used as personal test-dataset. Face detection performance of our proposed algorithm is analyzed in terms of Recall and Precision. As shown in Table 2, competitive performance of face-detection has been obtained by our scheme with short delays, which makes the proposed scheme more suitable for online and real-time applications. The effective use of GMC-maximization due to our algorithm for effective foreground-detection, has highly increased face-detection accuracy to reach 90% in high-resolution images. At the same time, we have avoided the increase of the computational-cost caused by GMC-maximization by focusing on a very small portion (block) of the image. The block in which the algorithm has efficient detection is smaller than 10% of the whole image size.

Herein, we have proved the efficiency of using our simple features for detecting both human blobs and faces presented in the detected-blobs with short delays and high accuracy. In the next Section 4.2.2, we will prove the efficiency of abstracting discriminative vectors of detected-faces in term of energy-consumption.

4.2.2. Energy and traffic efficiency analysis of face-detection algorithm at VS

For evaluating the energy-efficiency of our scheme, bus-station dataset has been performed using OMNET++/Castalia framework.

Table 3
Internal parameters of each visual sensor VS in the network.

Parameter	Description	Value
AoV	Angle of view of each camera	76°
DoV	Depth of view of each camera	30 m
C.Rate	Frames capture rate	0.58 fps
Vs	Supply voltage	4 V
CEPIC	Consumed energy per image capture	800 mJ
CEPIP	Consumed energy per image processing	277 mJ
DCRA	Drawn current while Radio is active (Tr | Rs)	46 mA
DCCA	Drawn current while camera is active	50 mA

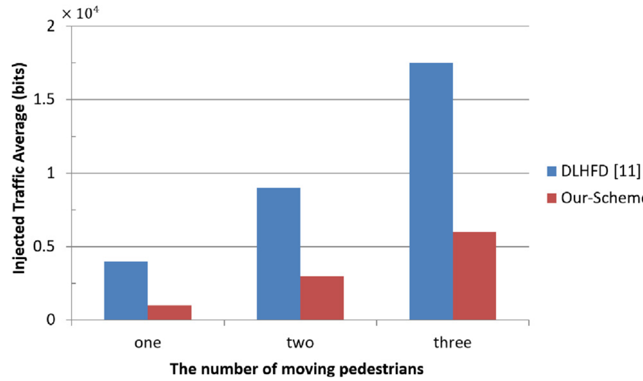

Fig. 9. Comparison results of average injected-traffic to network.

The parameters of this framework are set for each VS as presented in Table 3 and for network-distribution and communications as presented in Table 4.

For communications-cost analysis of our proposed scheme, we start with the evaluation of the injected traffic to the network. We have compared the injected-traffic average from each VS in our scheme with the results of the approach [11], which has been described in Section 2. As expected, transmission of discriminative vectors due to our approach has led to significant reduction of network-traffic compared to the approaches depending on transmission of raw pixel-level data of faces. Separated frames of bus-station sequence (test-samples) are shown in Fig. 8 and comparison results are shown in Fig. 9.

Considering the parameters presented in Table 3 and Table 4 and using 100-frames sequence of the pedestrian dataset as test-data, we have calculated the average in-node and in-network energy consumption of our scheme and compared our results with the results achieved in [10]. For energy consumption calculation, we have considered the following modes as described in our previous work [12]: 1 — Sleep mode: the current drawn by each VS in this mode is 9 mA. 2 — Active mode: the energy consumption of each VS in this mode is calculated due to the following formula:

$$E_{Consumed/Active} \approx \{C.Rate \times (CEPIC + CEPIP) + Vs \times (DCCA + DCRA)\} \times T_{mode(Active)} \quad (6)$$

Table 2
Experimental results of face-detection accuracy.

Data-Set	Recall	Precision	Average Delay-of-detection for each frame
Pedestrians (backdoor)	64%	69%	54 ms
Pedestrians (bus-station)	68%	76%	62 ms
Personal (1920×1088P)	82%	90%	165 ms

Table 4
Network parameters and communication protocols.

Parameter	Model/Value
Communication-protocol	X-Bee
Radio Model	IEEE 802.15.4, non-beacon CSMA mode
Routing Protocol	On-Demand (GPSR)
packet overhead	9 ms
Communication Range	35 m
Network Size	100 × 100 m
# Visual Sensors (VS)	15
# Scalar Sensor Nodes (SN)	30

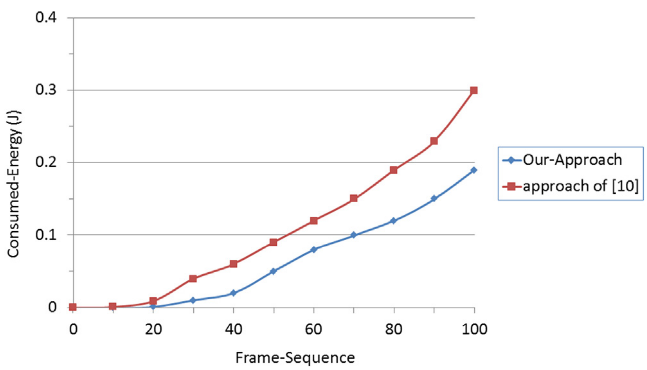

Fig. 10. Comparison results of average in-node energy-consumption.

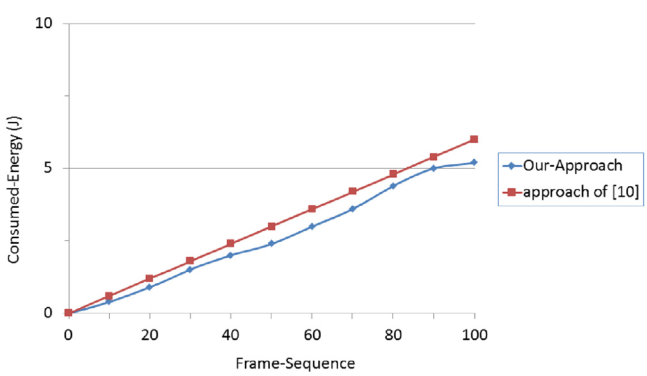

Fig. 11. Comparison results of in-network energy-consumption.

$T_{mode(Active)}$ is the time expended by VS in the active mode. CEPIC, CEPIP, Vs, DCCA and DCRA are defined in Table 3. As expected, our scheme overcomes the approach of [10] (which depends on transmitting the raw data of changed-blocks) in term of energy conservation. Comparison results of the average in-node energy consumption are shown in Fig. 10, whereas the comparison results of in-network energy consumption are shown in Fig. 11.

In-node energy consumption indicates to the average energy consumed by each VS for local-processing. Since initial frames are background, there is no triggered-VS at the beginning of the simulation-round and accordingly, in-node energy consumption is zero for the initial frames of the sequence as shown in Fig. 10. By receiving the following frames (containing moving-objects), the number of triggered VSs and accordingly the average in-node energy consumption will be increased. Since there is no feature-extraction process in [10], average in-node energy conservation of our approach has not shown big improvement compared to [10]. However, our novel vector-extraction approach has greatly reduced in-network energy consumption due to transmission of vectors instead of raw data and at the same time, has improved both detection and recognition accuracy. In-network energy-consumption is the sum of the energy

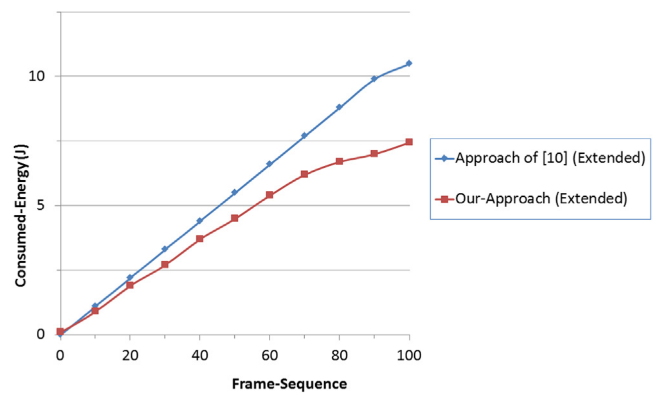

Fig. 12. Comparison results of in-network energy-consumption for extended network (200×100 m).

consumed by all VSN-members (VSs and SNs) due to transmitting, sensing and forwarding the information through the network and to sink. Accordingly, in-network energy consumption starts increasing by the first frame (even if being background) as shown in Fig. 11 indicating to the energy being consumed for required initial communications and network setup.

For evaluating the scalability of the proposed scheme in terms of network-size and the number of nodes in the network, new simulation sessions have been performed using the same parameters presented in Table 3 and Table 4 with extending the size of the network-area to be (200×100 m) employing 30-VSs and 60-SNs. As shown in Fig. 12, our scheme is performing better compared to the approach presented in [10] in term of in-network energy consumption by duplicating the network size and the number of nodes. By extending the size of the network, the number of communications between nodes would be increased and accordingly, the effect of reducing communication-cost on energy consumption appears more clearly.

As we have mentioned, data transmission from a visual sensor consumes more than doubled amount of energy, which is consumed by local processing of the same data [25,26]. Accordingly, by extending the network size, nodes consume more energy for transmitting data to longer distances and by increasing the number of VSs, the number of communications between them will be increased and the profit of transmitting vectors due to our approach will be more confirmed. We have increased the size of the network to (200×200 m) employing 60-VSs and 120-SNs using the same parameters and as expected, we have achieved better results compared to [10] as shown in Fig. 13.

In extended networks, our proposed scheme has achieved great improvements. As we have mentioned, in-node energy consumption is relative to local-processing at each VS, whereas in-network energy consumption is related to data transmission and receiving between VS nodes, which is highly affected by the area-size and the number of VSs in the network.

Table 5
Comparison results of in-node, in-network and total energy-saving for different sizes of the network.

Approach	Avg. In-node Energy consumption (J)	Energy saving percentages (in-node)	Total in-nodes energy-consumption (J)			In-network energy-consumption (J)			Energy saving percentages (in-network)			Total energy consumption (J)			Total energy saving		
			15-VS	30-VS	60-VS	15-VS	30-VS	60-VS	15-VS	30-VS	60-VS	15-VS	30-VS	60-VS	15-VS	30-VS	60-VS
Direct approach	≈1	Reference (0%)	≈15	≈30	≈60	≈40	≈55	≈100	Reference (0%)			≈55	≈85	≈160	Reference (0%)		
Approach of [10]	0.31	69%	4.7	9.3	18.6	6.2	10.7	19.7	84%	81%	80%	10.9	20	38.3	**80%**	**77%**	**76%**
Approach of [13]	0.41	59%	6.15	12.3	24.6	4.5	8.6	15.9	89%	84%	84%	10.6	20.9	40.5	**81%**	**76%**	**75%**
Approach of [3]	0.36	64%	5.4	10.8	21.6	7.3	11.2	21	82%	80%	79%	12.7	22	42.6	**77%**	**74%**	**73%**
Proposed approach	**0.18**	**82%**	**2.7**	**5.4**	**10.8**	**5.1**	**7.3**	**11.1**	**87%**	**87%**	**89%**	**7.8**	**12.7**	**21.9**	**86%**	**85%**	**87%**

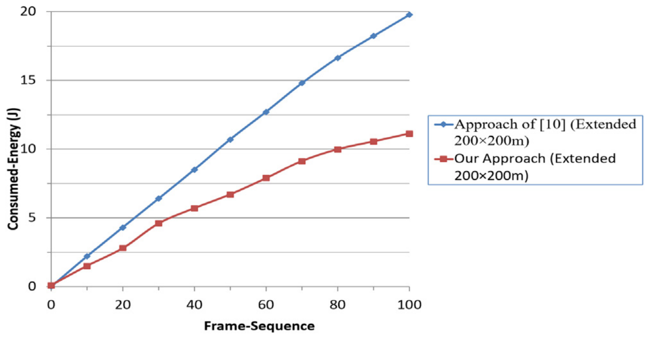

Fig. 13. Comparison results of in-network energy-consumption for extended network (200×200 m).

Although in-node and in-network energy consumptions are accumulatively calculated as shown in Fig. 10 to Fig. 13, energy saving computation is used for steady-state analysis of the scheme. Energy saving is computed at the end of each simulation-round (when the last frame of the sequence has been treated by the network). Table 5 summarizes the results of energy-savings for different sizes of the network (15-VS in 100×100 m, 30-VS in 200×100 m and 60-VS in 200×200 m).

In the direct-approach of object-detection based on VSN, there is no sleep-mode and the raw data of the captured images should be sent to sink when detecting an object appearance. Without any intelligence, the direct-approach is considered as a reference to which energy-saving of other approaches may be compared. In Table 5, we have compared energy-saving of our proposed scheme with the results of related works using the direct-approach as a comparison reference to show the impact of using intelligent local-processing algorithms on both in-node and in-network power consumption.

The improvement in energy-saving depends on the proportion of the burden, which is given to local processing at VS and on the efficiency of the local-processing algorithm being used. Approach of [13], which locally carries out detection and classification tasks and sends only reports of the results to sink, has achieved good in-network energy-saving but in return, it imposes relatively high average in-node energy consumption. Our proposed scheme uses a simple feature (W/H Ratio) for human-detection and searches only small portion of the image for face-detection and accordingly, achieves the best average in-node energy-saving among the related works. The proposed scheme also uses the already extracted features from each detected-face to concatenate a lightweight unique-vector to be sent to sink with low communication-cost and hence, provides the best improvement in in-network energy saving compared to other approaches.

The total energy-saving indicates to the overall improvement in energy-saving of each approach. The total energy-consumption is the sum of the total in-nodes and in-network energy consumption. Total in-nodes energy-consumption is the sum of the energy consumed by all VSs for local-processing after one simulation-round. Total energy-saving percentage has been computed according to the overall energy consumption.

As can be noticed from Table 5, the effect of average in-node energy-consumption on the total energy-consumption is related to the number of VSs in the network. Accordingly, approach of [10] provides slightly better total energy-saving compared to [13] in large network (60-VS in 200×200 of area). But, in return, approach of [13] has performed slightly better for smaller network (15-VS in 100×100 of area) in term of total energy-saving. Since we have achieved improvements in both in-node and in-network energy-savings, our proposed scheme provides significant improvement in the total energy-saving regardless of the number of VSs and the area-size of the network. We have achieved an average total energy-saving of 86% compared to the direct approach.

4.2.3. Experimental results of the face-recognition algorithm at sink

In this subsection, the performance of the recognition algorithm at sink would be analyzed. We have utilized Yale and FERET databases using OpenCv library with the parameters described in Section 4.1 for face recognition considering voting system for classification. Yale database contains 11-images for each individual. We have randomly chosen 4-images (faces) from each one of the 15-individuals as test samples. The discriminative vector extracted from each 4-images-sequence is considered as one test sample and the remaining 7-images of each individual are used as training samples. The implementations have been repeated 25-times and the average-results have been computed. Notice: for precise evaluation, discriminative vector extraction of the test samples considers the conformist parameters of HOG and PCA as described in Section 4.2. Classification results are presented in Table 6 in term of the correct recognition rate (CRR) and compared to one of the most recent works [27], which uses supervised approach for face recognition. In this approach, two classifiers (K-nearest neighbor (KNN) and SVM) have been employed over kernel discriminated analysis (KDA) features. As presented in Table 6, our approach obtains nearby results of accuracy but with significant reduction of computational-cost and delay due to our fast test approach.

Experiments are also performed on difficult datasets including extended Yale-B (2414 images of 38 individuals) in which, the images are exposed to extreme illumination changes. And FERET dataset in which, images are collected in semi-controlled environment and contains a total of 14126 images for 1199 individuals. The average results of these experiments are compared to face recognition approach provided in [28]. As presented in Table 7, our recognition approach, which depends on extraction of a unique discriminative-vector from multiple test-samples, obtains higher recognition-accuracy. Although using multi-scales of each training-sample may increase the time required for training the classifier, this training procedure is accomplished offline before initiating the operation of face detection. Since the recognition algorithm is accomplished at sink with rich-resources, recognition-accuracy and recognition-delay are the most important factors. Our approach has achieved high accuracy of recognition (CRR) with very short delays.

5. Conclusion

In this paper, we present a robust and energy-efficient face detection and recognition scheme, which is useful for many applications, including security and surveillance in VSN. The main idea of the proposed scheme is to improve local-processing of multimedia data at VS for the sake of reducing the communications-cost of the network. This improvement has been achieved by extracting light-weight but discriminative vectors from detected-faces to be sent to sink with low communication-cost. The results of many experimental sessions, performed on standard datasets and on personal images, prove that our proposed scheme has three important advantages: 1. Accurate and fast, which makes it suitable for real-time applications. 2. Energy-efficient and bandwidth-conservative, which makes it suitable for wireless resources-constrained environments. 3. Secure, due to using a novel dimensional reduction and data fusion approach for vector-extraction from images-sequence to be communicated between nodes. In the future, we will continue working on the efficiency of VSN trying different approaches such as using on-demand data transmission from VS to sink based on the request of the sink.

Table 6
Comparison results of face-recognition performance.

Dataset	Training samples	Approach	CRR	Delay-of-features extraction (test)	Delay of classification (prediction)
Yale	7 × 15 = 105	Our scheme	92.73%	0.27 ms	1.15 ms
		Approach of [27]	95.25%	0.32 ms	2.6 ms

Table 7
Comparison results of correct recognition rate (CRR).

Dataset	Training samples	Approach	CRR
EX-Yale-B	10 × 20 = 200	Our scheme	81.12%
		Approach of [28]	79.61%
FERET	5 × 20 = 100	Our scheme	93.32%
		Approach of [28]	91.44%

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2019.106014.

CRediT authorship contribution statement

Abdulaziz Zam: Conceptualization, Data curation, Methodology, Software, Writing - original draft. **Mohammad Reza Khayyambashi:** Supervision. **Ali Bohlooli:** Writing - review and editing, Supervision, Validation.

References

[1] C.B. Margi, et al., Energy consumption tradeoffs in visual sensor networks, in: Semantic Scholar, 2006.
[2] S. Aziz, D. Pham, Energy efficient image transmission in WMSN, IEEE Commun. Lett. 17 (6) (2013) http://dx.doi.org/10.1109/LCOMM.2013.050313.121933.
[3] Hadi S. Aghdasi, Shamim Yousefi, Enhancing lifetime of visual sensor networks with a preprocessing-based multi-face detection method, Wirel. Netw. 24 (6) (2018) http://dx.doi.org/10.1007/s11276-017-1451-z.
[4] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb.
[5] Z. Chen, Object tracking over a multiple-camera network, in: IEEE International Conference on Multimedia Big Data, 2015, http://dx.doi.org/10.1109/BigMM.2015.53.
[6] X. Wang, Intelligent multi-camera video surveillance: A review, Pattern Recognit. Lett. 34 (1) (2013) http://dx.doi.org/10.1016/j.patrec.2012.07.005.
[7] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference, Computer Vision and Pattern Recognition, CVPR, 2005.
[8] Z. Zivkovic, Improved adaptive gaussian mixture model for background subtraction, in: International Conference on Pattern Recognition, ICPR, 2004.
[9] X. Li, L. Wang, E. Sung, Adaboost with SVM-based component classifiers, Eng. Appl. Artif. Intell. 21 (5) (2008) http://dx.doi.org/10.1016/j.engappai.2007.07.001.
[10] Y.A. Ur Rehman, et al., A novel energy efficient object detection and image transmission approach for wireless multimedia sensor networks, IEEE Sensors J. 16 (15) (2016) http://dx.doi.org/10.1109/JSEN.2016.2574989.
[11] S. Najjar-Ghabel, Sh. Yousefi, Enhancing performance of face detection in visual sensor networks with a dynamic-based approach, Wirel. Pers. Commun. 97 (4) (2017) http://dx.doi.org/10.1007/s11277-017-4832-9.
[12] A. Zam, M.R. Khayyambashi, A. Bohlooli, Energy-aware strategy for collaborative target-detection in wireless multimedia sensor network, Multimedia Tools Appl. 78 (13) (2019) http://dx.doi.org/10.1007/s11042-019-7204-5.
[13] M. Civelek, A. Yazici, Automated moving object classification in wireless multimedia sensor networks, IEEE Sens. J. 17 (4) (2017) http://dx.doi.org/10.1109/JSEN.2016.2638853.
[14] Xin Wei, et al., A multi-scale method for HOG-based face recognition, in: International Conference on Intelligent Robotics and Applications, vol. 49, Springer, 2015, http://dx.doi.org/10.1007/978-3-319-22879-2.
[15] J. Yang, D. Zhang, A.F. Frangi, J.-Y. Yang, Two-dimensional pca: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) http://dx.doi.org/10.1109/TPAMI.2004.1261097.
[16] Shang-Hung Lin, An introduction to face recognition technology, Inf. Sci. Int. J. Emerg. Trans. Discipl. (2000).
[17] O. Demigha, W. Hidouci, T. Ahmed, On energy efficiency in collaborative target tracking in wireless sensor network: A review, IEEE Commun. Surv. Tutor. 15 (3) (2013) http://dx.doi.org/10.1109/SURV.2012.042512.00030.
[18] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The FERET evaluation methodology for face-recognition algorithms, in: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2002, http://dx.doi.org/10.1109/CVPR.1997.609311.
[19] Yale face database, 2009, http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
[20] Y. Wang, P.-M. Jodoin, et al., An expanded change detection benchmark dataset, in: Proc. IEEE Workshop on Change Detection at CVPR-2014, 2014, pp. 387-394.
[21] Pierluigi Carcagnì, et al., Facial expression recognition and histograms of oriented gradients: a comprehensive study, Plus 4 (2015) http://dx.doi.org/10.1186/s40064-015-1427-3.
[22] P. Dollar, et al., Fast feature pyramids for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 36 (8) (2014).
[23] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2) (2002).
[24] http://cpham.perso.univ-pau.fr/WSN-MODEL/wvsn-castalia.
[25] M. Nikzad, A. Bohlooli, K. Jamshidi, Video quality analysis of distributed video coding in wireless multimedia sensor networks, Int. J. Inf. Technol. Comput. Sci. 7 (1) (2015) http://dx.doi.org/10.5815/ijitcs.2015.01.02.
[26] M. Nikzad, A. Bohlooli, K. Jamshidi, Performance evaluation of error control schemes for distributed video coding over wireless multimedia sensor networks, Multimedia Tools Appl. 77 (15) (2018) http://dx.doi.org/10.1007/s11042-017-5397-z.
[27] M.Z.N. Al-Dabagh, et al., Face recognition system based on kernel discriminant analysis, k-nearest neighbor and support vector machine, Int. J. Res. Eng. (2018).
[28] Z. Chen, et al., Towards a face recognition method based on uncorrelated discriminant sparse preserving projection, Multimedia Tools Appl. 76 (17) (2017) http://dx.doi.org/10.1007/s11042-015-2882-0.

Abdulaziz Zam is currently a Ph.D. Candidate in the faculty of computer engineering, University of Isfahan. He received his B.Sc. and M.Sc. degrees in computer science from the University of Isfahan, IRAN in 2008 and 2013, respectively. His current research interests include Wireless sensor networks, Multimedia communications, Image processing and Computer vision.

Mohammad Reza Khayyambashi was born in Isfahan, Iran in 1961. He received his B.Sc. degree in Computer Hardware Engineering from Tehran University, Tehran, Iran in 1987. He received his M.Sc. in Computer Architecture from Sharif University of Technology (SUT), Tehran, Iran in 1990. He got his Ph.D. in Computer Engineering, Distributed Systems from University of Newcastle upon Tyne, Newcastle upon Tyne, England in 2006, he was research assistant during his Ph.D. course at University of Newcastle upon Tyne, Newcastle upon Tyne, England. He is now working as an Associate

Professor at the Department of Computer Architecture, Faculty of Computing Engineering, University of Isfahan, Isfahan, Iran. His research interests include Distributed Systems, Computer Networking, Mobile Cloud, Cloud Computing, Software Define Network (SDN), Mobile and Social Networks.

Ali Bohlooli received his BSc and MSc degrees in Computer engineering (with honors) from the Department of Electrical & Computer Engineering, Isfahan University of Technology, Iran, in 2001 and 2003, respectively. Subsequently, he received his Ph.D. degree in computer engineering from the University of Isfahan, Iran, in 2011. Presently, he is an assistant professor in the Faculty of Computer Engineering, University of Isfahan, Iran. His research interests include wireless networks and network modeling.