IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 December 2022, accepted 22 December 2022, date of publication 26 December 2022,
date of current version 5 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3232393

RESEARCH ARTICLE

A Distributed Rate-Control Approach to Reduce Communication Burdens in VSNs

SALMA ELEUCH[1,2], NADIA KHOUJA[1], SIMONE MILANI[2], (Member, IEEE), TOMASO ERSEGHE[2], AND FETHI TLILI[1]

[1]GRESCOM Laboratory, Higher School of Communications of Tunis (SUPCOM), University of Carthage, El Ghazala Ariana 2083, Tunisia
[2]Dipartimento di Ingeneria dell'Informazione, Università degli studi di Padova, 35131 Padua, Italy

Corresponding author: Salma Eleuch (salma.elleuch@supcom.tn)

ABSTRACT In visual sensor networks, the analyze-then-compress paradigm, where each camera process data and extract local features, is proved to be an efficient approach to reduce the amount of transmitted information. The bitrate can be further reduced by efficiently compressing the extracted features using a distributed feature coding technique. However, since the rate control is performed at the decoder, an abundant use of the feedback channel is needed to adjust the coding rate. Moreover, transmitting all extracted features, including irrelevant ones with no further contribution to the application accuracy, overloads the network. In this paper, we propose a novel feature selection and distributed coding rate control strategies that cope with these issues. The proposed strategies are designed to significantly reduce the transmitted bitrate and the communication burden with the sink, which implicitly reduces the energy consumption and the decoding delay. We show that, wisely selecting at the camera sensors level only the features effectively contributing to the application accuracy reduces the amount of transmitted information up to 34% while preserving accuracy. Furthermore, the cameras can collaborate periodically, by exchanging small amount of information about their selected features, to estimate the minimum transmission rate required for each feature based on a linear fitting model that takes into consideration the inter-camera correlation and the channel conditions. Significant average bitrate savings, reaching up to 37.71%, are achieved.

INDEX TERMS Distributed feature coding, rate control, feature selection, multi-view, visual sensor networks.

I. INTRODUCTION

In the last few years, Visual Sensor Networks (VSNs) have emerged as a potential enabler for a new class of applications in which vision is a key component, such as video surveillance, traffic monitoring, and many others [1], [2], [3]. Most of these applications require the processing and the transmission of huge amounts of data leading to high bandwidth usage and tremendous energy consumption. Satisfying these requirements is challenging due to VSNs' limited computational, communication and energy resources. Thus, many researchers have focused on finding efficient solutions for compressing, processing and transmitting the visual data in order to optimise the use of the network resources [4], [5], [6], [7], [8], [9], [10]. A widely used solution in this sense is the Analyze-Then-Compress (ATC) approach [11], [12], [13], [14], according to which the camera nodes perform some local processing to extract, encode and transmit only visual features which succinctly represent the most-informative parts of the image, thus using much less network resources than the pixel-level representation. The ATC paradigm constitutes a powerful solution for the applications that depend only on the results of the visual analysis.

Often, many cameras in a VSN capture the same scene or object from different angles resulting in some overlapped fields of view (FoV). Exploiting the high correlation between features extracted from these overlapped views can help further reducing the bit-stream flow, maximizing the coding efficiency and improving the analysis accuracy. In fact, using redundant information from several cameras can solve practical problems such as occlusions, illumination and pose variations. Features can be compressed (coded) by

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

exploiting the inter-view correlation through Distributed Feature Coding (DFC) for independent and low-complexity coding and joint decoding of the extracted features [4], [13]. Under the DFC approach, the inter-view correlation is exploited at the decoder side by using previously-decoded descriptors from a generic view (either spatially or temporally adjacent) as a Side Information (SI) to decode the currently-received information. An alternative method to exploit inter-view coding suggests a collaboration between neighboring cameras by exchanging their set of extracted descriptors as proposed in [15] and [16]. In this case, only the residual part between descriptors extracted from a camera chosen as base view and a neighboring camera is coded and transmitted to the sink node. However, this method can be more energy demanding compared to the DFC approach since inter-camera communication consumes energy, especially in the case of continuous real-time tracking applications and depending on the number of extracted features per camera. In order to reduce communication between cameras, authors of [17] proposed an optimization framework that decides whether two cameras collaborate, based on a predicted multi-view feature coding compression efficiency parameter.

Depending on the image content, the number of extracted features can be huge (reaching thousands of features). In the multi-view scenario cameras capturing the same object will extract very similar descriptors. Transmitting all the extracted features from all cameras will overload the network especially if some of them represent just redundant information that consume the network resources without contributing to the final analysis accuracy [18]. Therefore, it is important to select which descriptors to encode and to transmit in order to save energy and reduce the overall bitrate. Meaningful works focused on the selection and multi-view distributed coding of local features in the literature [19], [20].

In the above context, this paper elaborates upon and substantially improves the system model for multi-view vehicle tracking at roundabouts in VSNs described in our previous work [21], to enhance the obtained results for bitrate savings. This is achieved by presenting two novel key techniques that are able to efficiently control the rate demand of the VSN, and implicitly its energy consumption and processing delay, while preserving accuracy. We rely on the observation that the number of extracted features per camera can be huge for some frames requiring significant communication resources that might not be afforded by the VSN, as previously mentioned. For this reason, in the system model described in [21], after capturing and processing images (i.e., background subtracting and vehicle classifying), an appropriate feature selection stage is performed to extract only features that represent moving vehicles. In this paper, we add a second *feature selection* stage, aiming at wisely discarding those features belonging to the detected vehicles that consume the network resources without contributing to the analysis accuracy. Second, a *distributed rate-control* strategy for DFC is proposed where cameras collaborate periodically by exchanging negligible information about their selected features in order to better estimate the correct minimum required rate without the intervention of the decoder, thus efficiently shifting some of the coding rate control to the camera nodes. Note that, one key aspect of the proposal of this paper relies on the identification of a distributed algorithm. To this aim, we assume that the VSN can be divided into small *cameras cluster networks* and a relay network that links all the clusters to the sink node through multi-hop communications. A cluster network groups a number of smart camera sensor nodes capturing the same scene from different angles and capable of communicating among each other directly when it is needed. Each cluster network is identified by a central node related to the relay network, for control purposes, it is able to locally exchange data, and as such it is the basis of the distributed approach. Especially in those applications where the sink is faraway from the camera network (e.g., in swarms of drones), distributed feature selection and rate-control are of paramount importance in reducing the amount of data transferred to the sink node (energy savings), as well as in reducing the number of exchanges with the sink (processing delay mitigation) as decisions are taken locally and not centrally.

The main contributions of the present paper can be summarized as follows:

1) This paper aims at reducing the transmission bitrate by minimizing the total number of features to transmit over the VSN. To achieve that, we introduce a robust feature selection framework that aims at selecting the most relevant features to transmit, through a selection process that takes into account the contribution to the perceived quality at the application level (i.e., multi-view matching and tracking). Unlike the state-of-the-art solutions [19], [20] that mainly rely on performing feature selection at the sink node after receiving some information about the extracted features, in our solution, the feature selection is performed locally at the camera nodes without any information exchange.
2) A novel feature compression strategy is also designed to ensure enhanced bitrate reduction and robustness under severe communication conditions, e.g., low bandwidth links, bottlenecks, and limited energy. In fact, by periodically exchanging negligible information about their selected features, cameras in a cluster network collaborate to estimate the right source rate at which the features should be transmitted, given their correlation across cameras and the channel status, guarantying a successful descriptor reconstruction at the sink node. Unlike many of the solutions currently available in the literature where the sink is the only responsible for rate control by asking for more or less parity bits from cameras through a feedback channel [13], [20], [22], [23], [24], our system shifts some of the rate control to the cameras in order to reduce the abundant use of the feedback channel which implicitly reduces the processing delays.

The rest of the paper is organized as follows. An overview of the proposed system model is available in Section II.

Section III presents the proposed feature selection framework while Section IV presents the new coding strategy for DFC. In Section V, a number of experimental results that validate the novel techniques proposed in this paper by suitably measuring the system performance, i.e., bitrate reduction capabilities and feature matching accuracy, in a practical traffic monitoring scenario at roundabouts. Conclusions are finally drawn in Section VI.

II. SYSTEM MODEL OVERVIEW

We consider a VSN connecting multiple camera cluster networks to a sink node through a relay network with low bandwidth links, for the purpose of multi-view vehicles tracking. The overall system model used within each camera cluster network elaborates upon the one depicted in details in [21] (to which the interested reader is referred) and briefly summarized in Section II-A, with the addition of a new rate control strategy detailed in Section II-B.

A. MULTI-VIEW VEHICLE TRACKING SYSTEM MODEL

In this section, we review the system model for multi-view vehicle tracking at roundabouts using VSNs proposed in [21]. The system employs an ATC paradigm, where each camera collects visual data, detects and classifies moving vehicles, and then extracts a set of local features representing these vehicles to be encoded and transmitted to the sink node. Features are encoded separately at each camera node using DFC, and then jointly decoded at the sink node by exploiting the inter-view correlation based on the approach presented in [13]. Once features are successfully retrieved at the sink node, one-view tracking and multi-view matching for switching the tracking from one view to the other are performed.

More specifically, it encompasses the following steps:

1) **Detecting moving vehicles**: First, background subtraction is applied for each acquired video frame to detect all moving objects using the Mixture of Gaussians (MoG) subtractor. Then some morphological transformations are added to the obtained binary image in order to refine the detected foreground objects. The first basic operation is dilation which helps in making the foreground object more visible by increasing the boundaries size. This step is essential to group close parts representing the same object. Then the dilated binary image is thresholded in such a way that the detected shadows are removed. Since the target application is to track vehicles, a classification step is performed to select from the moving objects only those representing vehicles. The YOLOv2 detector is used to accomplish this task after training it on a huge data set, termed SupCom Roundabout database, collected from camera sensors placed around a roundabout (for more details about the constructed dataset, see [25]). YOLOv2 is a real-time object detector and classifier based on 19 convolutional neural network layers.
2) **Feature extraction and clustering**: After detecting moving vehicles, each camera node extracts local features. The latter are a compact representation of the local content of an image patch that differs from its immediate surrounding by an image property. In this proposed system, Speeded-Up Robust Features (SURF) features [26] are extracted, which are robust features that are capable of representing the most salient characteristics of vehicles even in the presence of occlusion, illumination and pose change. Next, the extracted features are assembled into clusters by exploiting the `K-means` clustering algorithm that aims at grouping data points into K clusters by reducing within-cluster variances [27]. For the initialization process, the `K-means++` approach [28] is selected. Each cluster is identified by an ID and a centroid, in such a way that the features belonging to the same cluster have similar descriptors (i.e., they are correlated features). The set of clusters' IDs and centroids are computed offline and assumed to be a common knowledge between both the cameras and the sink so that the cluster assignment can be performed at each camera node without any exchange of information.
3) **DFC and transmission**: Features are separately encoded by a Slepian-Wolf (SW) encoder [29], which is built using a (6,4) regular systematic Low-Density Parity-Check (LDPC) encoder of rate 1/3, whose parity information bits as well as the cluster ID to which the encoded descriptor belongs are forwarded to the sink node. To construct the parity check matrix H, the predefined function `parity_check_matrix` from the `pyldpc` library in `python` is harnessed, which builds a regular Parity-Check Matrix following Gallager's algorithm [30].
4) **Data reconstruction**: The sink exploits the inter-view correlation to jointly decode the received features. The centroid is used for generating the Side Information (SI) needed in the estimate of the descriptor from the received parity bits; the SI can be the centroid itself, retrieved from the received cluster ID, or the already decoded descriptors from previous frames and from all cameras, belonging to that cluster (stored in a buffer at the decoder). A statistical Correlation Noise Model (CNM) is then used to estimate the correlation noise between the constructed SI and the true descriptor. From the estimated CNM, the Log-Likelihood Ratios (LLRs) are computed which are needed to run the Belief Propagation (BP) algorithm for the SW decoding. In case of decoding failure, the sink requests more parity bits from the camera node using the feedback channel.
5) **Multi-view vehicle tracking**: First, one-view feature-based tracking is performed using the tracker described in [25], where inter-frame feature matching is performed and features' trajectories are constructed by connecting all matched features over time. In case of occlusion, i.e., one-view tracking failure, the system switches and continues tracking the same object from

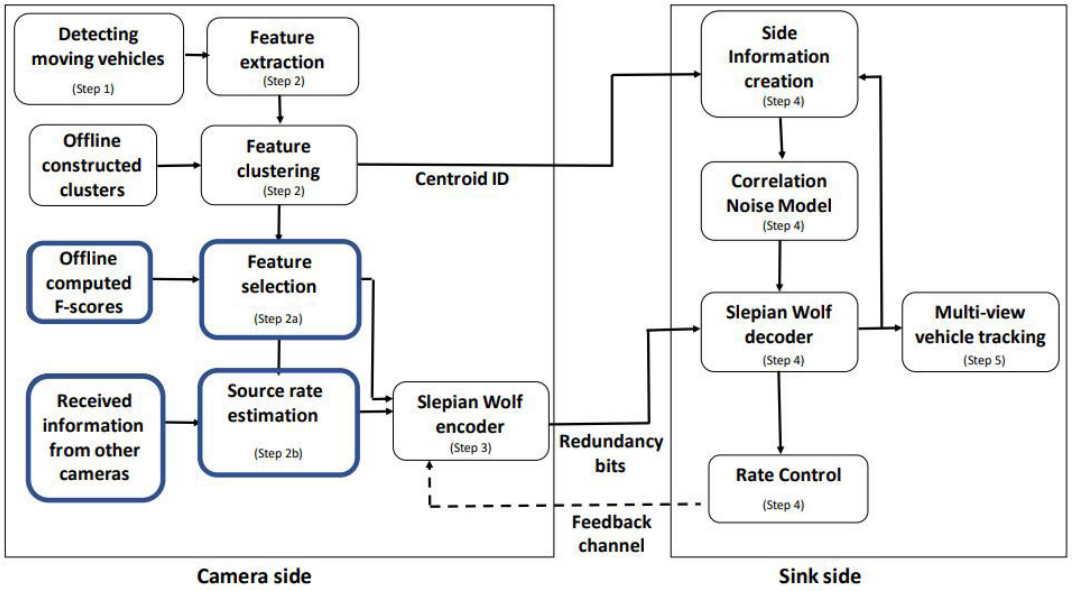


FIGURE 1. Pictorial representation of the overall proposed system.

another view with better sight of it. The identification of the same object in two different views is realized by applying multi-view features matching.

B. THE NEW PROPOSED SYSTEM MODEL

The described system in Section II-A is here optimized with two dedicated rate-control components, namely:

1) a robust feature selection strategy, and
2) an enhanced rate-control for the DFC,

both distributedly applied at a node level before encoding features, unlike state-of-the-art solutions that implement both actions at the sink side. As a proof-of-concept, the reliability and rate savings ensured by these two key components are tested in the VSN multi-view vehicle tracking scenario of [21]. Interestingly, although tested in a specific tracking context, the constituent idea is of general applicability in a wide range of VSNs applications such as assisting lonely persons in elderly care [31], tracking pedestrians [32], and wild fire detection for disaster management [33]. Figure 1 depicts the global system architecture for the proposed multi-view features coding and tracking as described in [21] with the addition of the new rate-control components.

The final system model proposed in this paper encompasses the steps described in Section. II-A with the extension of step 2 by the following sub-steps:

2a) **Feature selection**: The feature selection process is based on choosing only features with a matching accuracy metric greater than a certain threshold. For more details on the choice of the threshold and feature selection framework see later Section III.

2b) **Source rate estimation**: Each camera shares with neighbouring cameras some information about its newly extracted features (basically the cluster identifier and the 3D PCA components of the constructed descriptors) to estimate the inter-view correlation among features and exploit it to determine an accurate approximation of the source rate to attribute for each feature. The status of the transmission channel is also considered for the rate estimation. The detailed coding strategy proposed for the source rate estimation is further explained in Section IV.

The same SW decoder described in Section II-A step 4 is used to reconstruct the received vectors. However, with the new proposed coding strategy (i.e., source rate estimation), most of the rate control is performed at the encoder side. Therefore, a controlled and limited use of the feedback channel is ensured; if the decoding of the received feature fails, the sink can selectively request more parity bits from the cameras.

III. FEATURES SELECTION

A. RATIONALE

Feature selection consists in reducing the communication burden between cameras and the sink by choosing to transmit only relevant features that contribute to the final analysis accuracy. In this paper, we propose to evaluate the accuracy of multi-view feature matching through two metrics. The first metric is measuring the diversity of extracted features and prioritizing the most different features from what previously selected. The diversity metric limits the number of features per cluster each camera transmits while favoring a diverse selection of descriptors from different clusters (i.e., presenting different objects or different parts of the same object). For each cluster one feature is selected by choosing the feature with the highest hessian response.

The second metric is the F-score, namely the harmonic mean of *precision* and *recall* values of matching

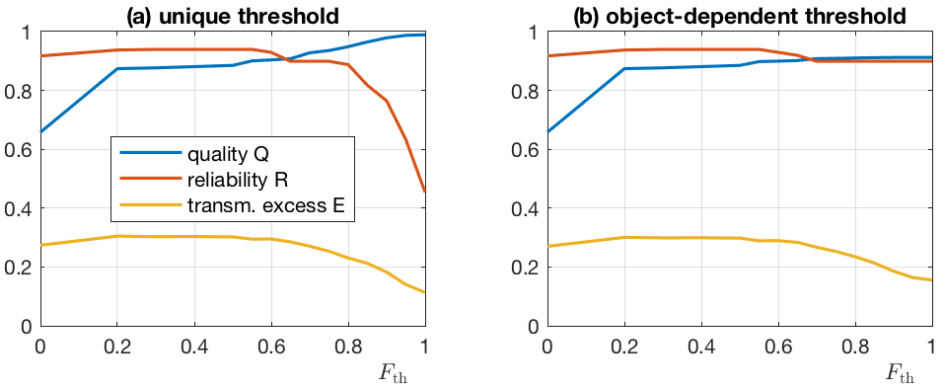


FIGURE 2. Key accuracy parameters versus F_{th} for (a) a unique threshold and (b) an object-dependent threshold.

(see also [21]), which is a crucial quality indicator in multi-view object recognition or object tracking applications. The F-score is assumed to belong to the range [0, 1], with 1 denoting perfect accuracy. The feature selection process is simply based on choosing features with an accuracy metric F-score greater than a certain threshold, and we envisage that the threshold value is identified by one of the following methods, namely:

a) *unique threshold*, a unique threshold F_{th} is identified for all features, or
b) *object-dependent threshold*, a different threshold is envisaged for each object, as they might carry different accuracies (e.g., object in the foreground, and object in the background partially occluded); the threshold is set to the minimum between a reference threshold F_{th} and the average value of the F-scores of the object's features.

Since features are assembled in clusters, the F-score of each feature is actually the F-score attributed to the cluster to which it belongs. Specifically, an offline multi-view matching of descriptors extracted from two cameras with overlapped FoVs, through 200 frames from the SupCom roundabout database described in [25], is performed and then an F-score is attributed to each cluster based on the results of matching features belonging to it.

Note that the feature selection is performed locally inside each cluster network based only on the clusters' attributed F-scores, which are assumed to be common knowledge between all cameras, thus entailing no information exchanges.

B. THRESHOLD CHOICE FOR F-SCORE BASED SELECTION

The F-score based feature selection process discards all features with F-score less than a certain threshold. If the threshold is sufficiently high, only relevant features will be transmitted, leading to very few wrong matches. However, if the threshold gets too high, the number of selected features decreases significantly. As a result, some important parts of the image or some objects of interest cannot be recognized or tracked anymore. Therefore, it is essential to choose the threshold F_{th} wisely. To do so, we identify three key accuracy parameters:

- a *quality function* Q that expresses the (weighted) average F-score of the selected features: the higher Q, the most accurate the selected features;
- a *reliability indicator* R, denoting the capability of the selection system in obtaining a minimum number of matched features per each object close to a certain target m_{ref}, specific to the chosen application: the closer R to 1, the most accurate the selection;
- a *transmission excess measure* E, expressing (in the average) the fraction of selected features in excess of the m_{ref} target: the closer E to 0, the most accurate the selection.

By denoting with $\mathcal{O}$ the set of objects, by $\mathcal{F}_o$ the set of features belonging to object o, by $|\cdot|$ the cardinality of a set, by F_n the F-score of the nth feature, by S_n the indicator function which is equal to 1 if the n^{th} feature is selected (i.e., if its F-score is greater than the threshold) and 0 otherwise, and by m_o the number of matching features of object o, then for an object $o \in \mathcal{O}$ the above key accuracy parameters can be expressed as

$$Q_o = \frac{\sum_{n=1}^{|\mathcal{F}_o|} F_n S_n}{\sum_{n=1}^{|\mathcal{F}_o|} S_n}, \; R_o = \frac{\min(m_o, m_{ref})}{m_{ref}}, \; E_o = \frac{\sum_{n=1}^{|\mathcal{F}_o|} S_n}{m_{ref}}, \tag{1}$$

while Q, R, and E are the corresponding averages over all the objects.

The behaviour of the three accuracy parameters for our tracking application is shown in Fig. 2 for (a) a unique threshold and (b) an object-dependent threshold, as a function of the F-threshold F_{th}, and for $m_{ref} = 5$. Curves were computed offline in order to identify a reasonable procedure for optimizing F_{th}. The chosen value F_{th} should maximize both the quality and the reliability functions, Q and R respectively, while minimizing the transmission excess E. Note from Fig. 2 (a) that, when F_{th} increases quality Q increases and the transmission excess E decreases, to the detriment of deteriorated

reliability values R. This practically sets the working point to $F_{th}^{*} \simeq 0.7$. Instead, as it can be seen from Fig. 2 (b), the object-dependent threshold significantly improves and stabilizes the behaviour of R, at the cost of a slight increase in E, and of a saturation of the quality parameter Q below one, yet still at a good accuracy value $Q \simeq 0.9$. Thanks to this stabilization effect, the object-dependent threshold is less dependent on F_{th}, and its performance is generally improved with respect to the working point choice $F_{th}^{*} \simeq 0.7$ in Fig. 2 (a), hence it is the approach to be preferred for reliability, performance, and robustness to the choice of F_{th}.

IV. DISTRIBUTED RATE-CONTROL STRATEGY

A. RATIONALE

As we anticipated in Section II, in the state-of-the-art approach [13], [21] features are assigned to clusters whose centroids are offline calculated and known throughout the entire network. However, due to the offline approach, features affected to the same cluster are likely to be weakly correlated, not strongly; as illustrated in Fig. 3, the cluster is usually a sparse collection of features, typically organized in small groups. For this reason, estimating the coding rate and constructing a SI for DFC decoding based on the centroid might be inefficient in some cases. In order to improve the encoding strategy, features in a cluster can be partitioned in *subgroups* of highly correlated features, as illustrated in Fig. 3. Features belonging to a subgroup can be transmitted at a lower rate if another feature of the same subgroup is used as a SI at the decoder. We therefore assume that the selected features are encoded following this strategy:

1) upon the creation of a new subgroup, only one feature per subgroup is transmitted at the *highest rate*, and is considered as the subgroup representative; the remaining features are transmitted at a *lower rate* so that they can be decoded by using the subgroup representative as SI;
2) once a subgroup is formed and some features belonging to it from the frame of its creation are already transmitted to the sink, then by assuming that at least one of these features is successfully decoded, all new selected features attributed to this subgroup are encoded with *lower rate*.

The subgrouping classification clearly aims at efficiently exploiting the correlation among features at both cameras and sink sides, thus implementing an efficient rate-control strategy and accurate SI reconstruction.

Operationally this requires that, in each camera cluster, cameras share periodically with neighbours some information about the newly selected features in order to identify, according to their correlation, subgroups of highly correlated features with distances between each other smaller than those separating them from the centroid. The number and identifiers of active subgroups are then updated, and a subgroup representative is elected. The information which needs to be exchanged is mainly the cluster identifier plus the 3D principal components of descriptors (using the Principal Component Analysis (PCA) method), which in total consists of a few bits of information, and which locally occupies a limited transmission bandwidth at the camera cluster level.

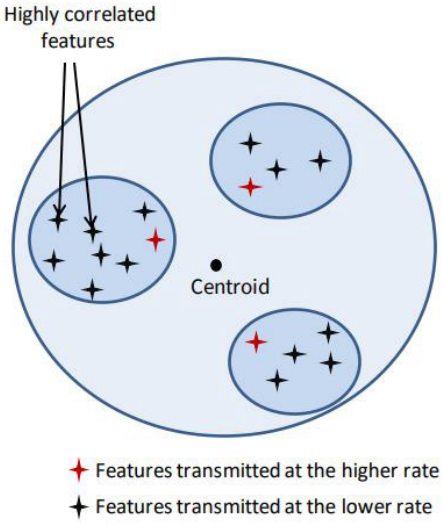


FIGURE 3. Subgrouping strategy within each cluster for a better modelization of the intra- and inter-view correlation.

To form subgroups, each camera estimates the inter-camera correlation using the received information from neighbours belonging to the same camera cluster network. First, the clusters to be divided into subgroups are identified using the received clusters' IDs. Then, for each identified cluster, the process of forming subgroups is realized by executing the following steps:

1) reordering the features from all cameras belonging to the cluster in a descending order with respect to their correlation with the centroid denoted by $corr(PCA_f, PCA_c)$, where PCA_f and PCA_c are the 3D PCA vectors representing the feature f and the centroid c, respectively.
2) considering the feature with the maximum correlation as a new subgroup center (first subgroup).
3) for the rest, comparing the correlation with the centroid (i.e., $corr(PCA_f, PCA_c)$) to the correlation with the already formed subgroups' centers (i.e., $corr(PCA_f, PCA_g)$, where g is the subgroup center). According to the maximum correlation value, the feature is whether affected to an existing subgroup or considered as a new subgroup center.

The resulting scheme is summarized in Algorithm 1.

B. PUNCTURING MODEL

The approach used to generate *higher or lower* rates is a puncturing approach where bits representing the entire parity part of features after SW encoding are, respectively, *lightly or strongly* punctured. For each feature $f_{i,n}$ selected by camera i at frame n, we denote by $\rho_{i,n}$ the fraction of punctured bits. The estimation of $\rho_{i,n}$ depends on the required reconstruction

Algorithm 1 Forming Subgroups Within a Cluster

1: Reorder the set of features $\mathscr{F}$ belonging to the cluster identified by the centroid c in a descending order.
2: Consider the first feature f_1 (i.e., most correlated with centroid c) as the first subgroup header.

$$\mathscr{S}_g \leftarrow \{f_1\}$$

▷ $\mathscr{S}_g$ is the set of the constructed subgroups' headers
3: **for** f in $\mathscr{F} - \{f_1\}$ **do**
4: Find maximum correlation between feature f and elements of the set $\{c, \mathscr{S}_g\}$ denoted by $MaxCorr$
5: **if** $MaxCorr = corr(PCA_f, PCA_c)$ **then**
6: Feature f is the header of a new subgroup

$$\mathscr{S}_g \leftarrow \{\mathscr{S}_g, f\}$$

7: **else**
8: Feature f is affected to the subgroup identified by $g^* = \arg\max_{g \in \mathscr{S}_g}(corr(PCA_f, PCA_g))$
9: **end if**
10: **end for**

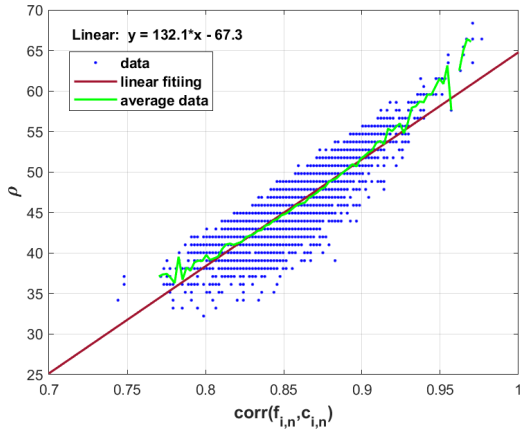
precision; the smaller the tolerable precision is, the bigger $\rho_{i,n}$ is and thus the lower is the transmission cost since less information is transmitted through the network. However, choosing a very high $\rho_{i,n}$ value (i.e., a very low tolerable precision) entails many decoding errors and a bad reconstruction accuracy. Therefore, $\rho_{i,n}$ must be wisely identified.

The fraction of punctured bits $\rho_{i,n}$ for the feature $f_{i,n}$ is chosen via

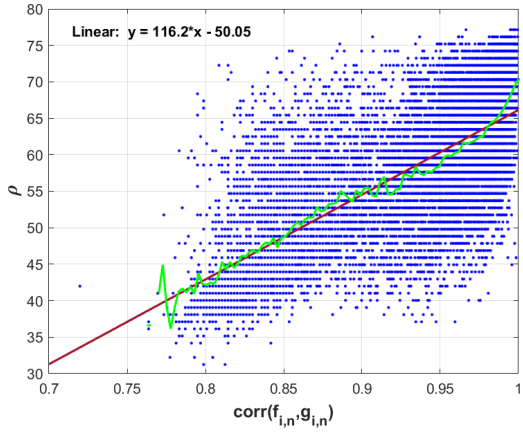
$$\rho_{i,n} = \begin{cases} \bar{\rho}_{high}(SNR, corr(f_{i,n}, c_{i,n})) & , \text{ high rate} \\ \bar{\rho}_{low}(SNR, corr(f_{i,n}, g_{i,n})) & , \text{ low rate} \end{cases}$$

where $\bar{\rho}_{high/low}$ is a mapping function that generally depends on the Signal to Noise Ratio (SNR) value, as well as on the correlation (corr) with a reference feature, i.e., the centroid $c_{i,n}$ for high rate coding and the subgroup's identifier $g_{i,n}$ (in its 3D PCA components form) for low rate coding.

Taking into consideration the aforementioned and assuming a transmission channel with an Additive White Gaussian Noise (AWGN), $\bar{\rho}_{high/low}$ is offline constructed by a *linear* surface fitting model. The fitting models were both computed offline for more than 15000 features extracted from vehicles detected in 200 frames and decoded successfully using the SW decoder as described in [21]. The fitted curves of $\bar{\rho}_{high/low}$ are illustrated in Fig. 4 for a fixed SNR= 10dB, while Fig. 5 depicts the bilinear fitting taking into account both correlation and SNR. The Root Mean Square Error (RMSE) values measuring the goodness of the linear fitting models when $SNR = 10$dB for $\bar{\rho}_{high}$ and $\bar{\rho}_{low}$ are $RMSE_{high} = 0.019$ and $RMSE_{low} = 0.072$ respectively. For the bilinear surface fitting, the RMSE values for high and low coding rates are respectively 0.059 and 0.119. Note that the RMSE values are slightly higher for low rate coding due to the fact that the estimated correlation in this case is determined on



(a) High rate coding



(b) Low rate coding

FIGURE 4. Linear fitting of $\rho_{i,n}$ (expressed in percentage) when SNR = 10dB and for: (a) high rate coding $\bar{\rho}_{high}$ (RMSE = 0.019), and (b) low rate coding $\bar{\rho}_{low}$ (RMSE = 0.072).

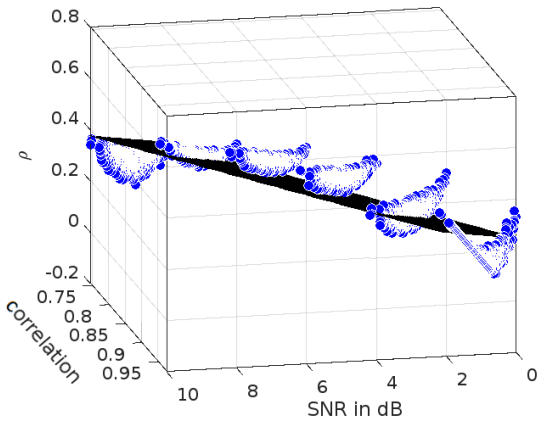
the 3D PCA components rather than the full 64 components descriptors.

A linear fitting model was chosen to estimate the puncturing fractions $\bar{\rho}_{high}$ and $\bar{\rho}_{low}$ mainly for two reasons. First of all, the linear model is simple yet efficient in fitting the data with 95% confidence bounds and small RMSE values (<0.1). Moreover, considering the average ρ values for each correlation level, the corresponding curves for both high and low coding cases (green curves in Fig. 4) are very close to the fitted curves (red curves). The second reason is that, considering more complex fitting models such as quadratic, exponential and power models, same behavior as the linear model is observed for fitting the majority of data points as it can be seen from Fig. 6

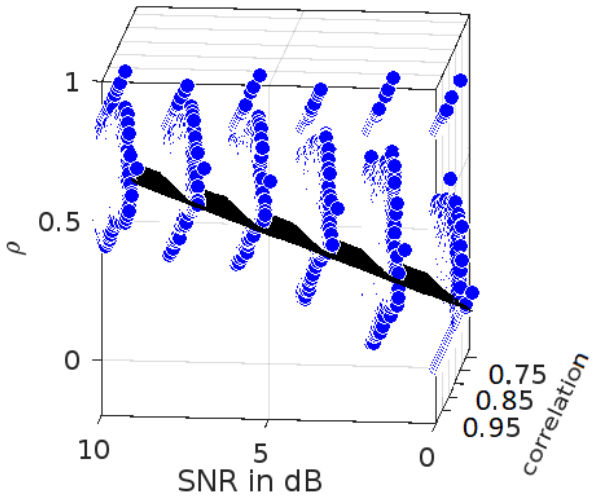
V. EXPERIMENTAL RESULTS

A. IMPACT OF FEATURE SELECTION AT THE APPLICATION LAYER

To evaluate the performance of the feature selection methods described in Section III-A, we investigate their ability to

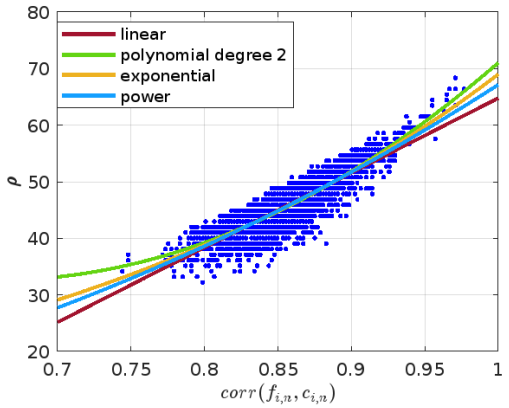


(a) High rate coding

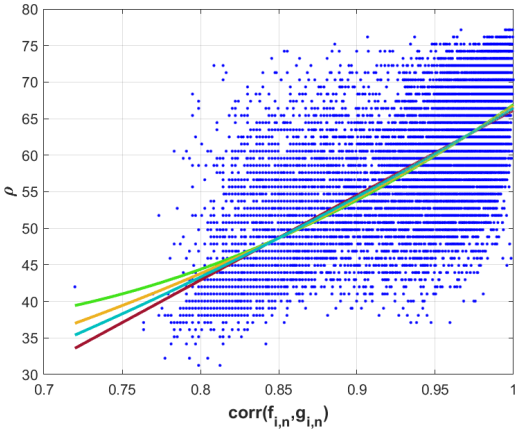


(b) Low rate coding

FIGURE 5. Bilinear surface fitting of $\rho_{i,n}$ for: (a) high rate coding $\bar{\rho}_{high}$ (RMSE = 0.059), and (b) low rate coding $\bar{\rho}_{low}$ (RMSE = 0.119).



(a) High rate coding



(b) Low rate coding

FIGURE 6. Comparison between different fitting models for: (a) high rate coding and (b) low rate coding.

reduce the amount of transmitted information as well as their impact on the multi-view matching accuracy. All the simulation results are obtained in the context of multi-view vehicle tracking application. The multi-view matching accuracy is evaluated according to the Receiver Operating Characteristic (ROC) curves of true versus false positive rates and the F-score metric. The results are computed in the case of ideal channel conditions and no bits puncturing at transmission to highlight the impact of feature selection on matching accuracy independently on the new coding strategy.

The total number of features to transmit after diversity-based selection is reduced by an average percentage of 5.41%. With the diversity-based approach, a diverse selection is guaranteed; each camera selects features belonging to different clusters. However, the achieved saving rate is considered very low. A much more significant saving is achieved with the chosen object-dependent F-score threshold described in Section III-B reaching 34%. In other words, 34% of the extracted features were considered irrelevant and discarded by the selection algorithm to save bitrate and transmission energy. Another possible approach is to combine the two metrics (i.e., F-score and diversity applied together) to select features. In this case, the number of transmitted features after selection is reduced by 45.53%. Even though the latter feature selection approach enables the highest saving rate, it restricts the reliability of the proposed coding strategy. In fact, the new distributed rate control coding strategy described in Section IV is mainly based on estimating the inter-camera correlation at camera side by forming subgroups. The formation of subgroups, and hence the accuracy of the correlation estimation, depend directly on the number of features extracted from different cameras belonging to the same cluster. By applying the diversity-based selection, the number of features belonging to the same cluster from all cameras is reduced.

To investigate further the efficiency of the proposed feature selection methods, their impact at the application layer is provided in Fig. 7 and Fig. 8. For the F-score based method, the chosen object-dependent threshold is used with a maximum limit of $F_{th} = 0.85$. The only purpose of setting

a constraint ($F_{th} = 0.85$) is to limit the maximum value the average F-score per object can take, so that a sufficient number of features per object transmitted and later matched is ensured. Note that this threshold value may change from one application to another depending on the analysis results. However, given the fact that the proposed feature selection method based on the average F-score per object stabilizes all key accuracy functions defined in Section III-B (see Fig. 2), not considering any threshold F_{th} will lead to almost the same results as those presented in Fig. 7 and Fig. 8.

Fig. 7 illustrates ROC curves of true versus false positive rates in the considered vehicle tracking scenario for the cases of no selection, F-score based selection, diversity-based selection, and hybrid selection combining the F-score and diversity metrics. The curves were obtained by varying a discrimination threshold d for classifying matches into true positive TP or false positive FP; the threshold d is the distance between two matched descriptors under which the match is considered correct. We consider a true positive TP as a correct matching of features belonging to the same vehicle, and a false positive FP as a matching between features belonging to different vehicles. On the other hand, the multi-view matching accuracy, measured by the F-score metric, for all aforementioned cases are shown in Fig. 8. The multi-view matching F-scores of the decoded descriptors are computed for each discrimination threshold d.

From Fig. 7 we can observe that the diversity-based selection method (yellow curve) has the best ROC curve, even better than the curve obtained where no selection is performed (blue curve). This can be explained by the fact that distinctiveness of features is a very important property for having accurate matching results. With the diversity-based selection, only distinctive features with high hessian response are selected. The matching F-scores achieved with the diversity method are very similar to those obtained when no selection is performed (see Fig. 8), especially at high discrimination thresholds ($d \geq 0.06$). The almost overall superposition of the two curves for F-score based selection and no selection cases in Fig. 7 and Fig. 8 confirms that, even though the number of transmitted features is reduced (bit-rate reduction), the matching accuracy is preserved and the proposed approach meets its target. Moreover, for high discrimination thresholds ($d \geq 0.06$), higher F-score values (around 0.9) are achieved with the proposed feature selection method compared to the cases of no selection performed and diversity-based selection, as it can be inferred from Fig. 8. With the final feature selection approach, i.e., hybrid selection, similar matching F-scores to those obtained in the case of F-score based selection are achieved. However, the performance of the ROC curve in the case of hybrid selection is slightly deteriorated compared to the other selection methods (see Fig. 7). This deterioration is due to the fact that a very high number of features are discarded by the hybrid selection method, leading to an insufficient number of features per object to be matched. This in return results, in some cases, in more wrong matches (either false positives or false negatives).

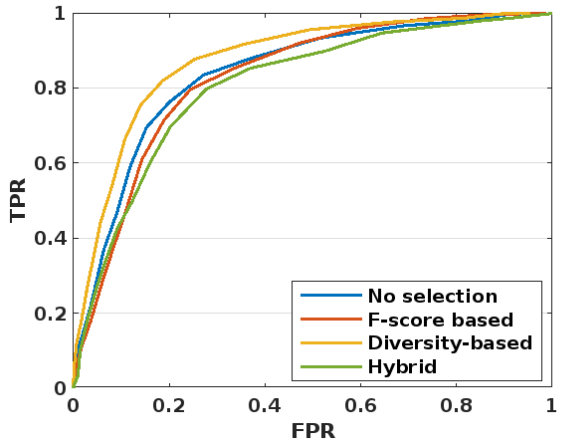


FIGURE 7. ROC curves with and without features selection.

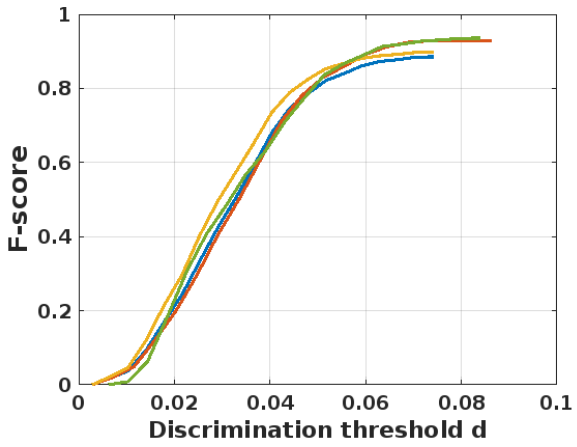


FIGURE 8. Matching accuracy metric F-score with and without features selection as a function of the discrimination threshold d.

Taking into consideration all the aforementioned, we chose to work with F-score based selection approach since a significant bitrate saving can be achieved while preserving a good matching accuracy compared to the case of no selection performed.

B. IMPACT OF THE NEW CODING STRATEGY ON THE BITRATE SAVINGS

To evaluate the efficiency of the proposed coding strategy, we provide in Fig. 9 the Frame Error Rate (FER) performance of the SW decoder with the new coding strategy (dashed line) against the standard strategy (straight lines), for different SNR values. In the subgroup coding strategy, for each feature $f_{i,n}$, a puncturing fraction $\rho_{i,n}$ is estimated using the fitting model described in Section IV-B. In the case of standard SW decoder, however, a constant puncturing value $\rho_{i,n} = \rho$ is applied to all features. Different ρ values ranging in percentage from 0 to 50% were used to compute the FER in the case of standard SW decoder, which are illustrated in Fig. 9.

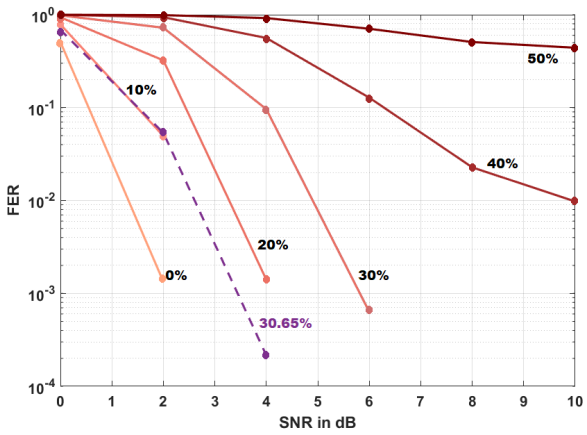


FIGURE 9. FER performance evaluation versus the SNR.

As it can be seen from Fig. 9, the proposed coding strategy performs similarly to the standard coding strategy with 10% bitrate saving, but guarantees much higher (average) puncturing percentages ranging from 16.15% at low SNRs (≤2dB) to 37.71% at slightly higher SNRs (≥4dB). The average saving achieved using the new coding strategy through all SNRs is 30.65%. Moreover, by comparing the FER results obtained when the new coding strategy is applied to those obtained for constant puncturing percentage $\rho = 30\%$, we can observe that much lower FER values are achieved even though the bitrate saving is almost the same in both cases. In other words, the proposed coding strategy in this paper entails less decoding failure of received descriptors, which implicitly results in less use of the feedback channel and less decoding delays.

The main idea of the new coding strategy proposed in Section. IV is to estimate the inter-view correlation at the camera side and exploit it to estimate the necessary puncturing fraction value for each selected feature. The inter-view correlation is estimated based on the subgroup formation after exchanging negligible amount of information among cameras. Therefore, to further analyze the accuracy of the proposed coding strategy, we evaluate the effect of the loss of some messages, due to the transmission channel noise, on the bitrate savings and application analysis performance. We denote by P_{loss} the probability of losing some messages during the information exchange process between cameras belonging to the same camera cluster network. We varied the value of P_{loss} for each SNR, and computed the average saving rate (i.e., average estimated puncturing fraction) as well as the achieved FER for 100 consecutive frames. Since perfect FER values (FER $= 0$) are achieved for high SNRs (SNR $\geq$ 6dB) when $P_{loss} = 0$ (see Fig. 9), we present the obtained results at SNR $=$ 6dB in Table. 1.

From Table. 1 we can observe that when $P_{loss} = 0.05$, the average saving rate has increased. This is expected considering that the lost messages were not taken into account for forming subgroups. Therefore, inexact inter-view correlation

TABLE 1. Evaluation of the bitrate savings and the FER for different P_{loss} values when SNR = 6dB.

P_{loss}	0	0.05	0.1	0.2	0.5
FER ($\times 10^{-4}$)	0	5.9	9	15	21
average savings in %	27.84	30.31	30.38	30.47	30.69

is estimated which impacts the estimation of the puncturing fraction. For some features, the estimated puncturing fraction is too high with respect to the maximum acceptable fraction (estimated when $P_{loss} = 0$), leading to some decoding errors. This also explains the slightly raised value of the FER. As the loss probability P_{loss} increases, the FER and the average saving rate slowly increase as well. In fact, even though more messages are lost, the number of features per cluster considered during the subgroups formation process is sufficient thanks to the redundant information extracted by cameras belonging to the same camera cluster network. In addition, the features that were not considered for forming subgroups because of the messages loss are automatically coded with respect to the cluster's centroid, which is still correlated enough for estimating the puncturing fraction. Interestingly, the tiny increase in the FER has a inconsiderable effect on the multi-view matching accuracy; same ROC curve performance obtained when $P_{loss} = 0$ (red curve in Fig. 7) is achieved.

VI. CONCLUSION

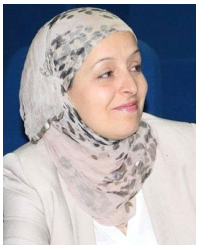
In this paper, we propose an improved distributed feature coding solution aiming at shifting the feature selection and some of the rate control to camera side in order to reduce the transmission bitrate and the decoding delay. Cameras select most relevant features based on feature matching accuracy computed offline cluster wise. Moreover, cameras collaborate periodically to estimate the exact source rate needed for each selected feature based on the inter-view correlation and the transmission channel conditions. The experimental results demonstrate that the amount of information to be transmitted to the sink can be reduced by 34% using the proposed feature selection algorithm. Furthermore, significant additional bitrate savings reaching 37.71% can be achieved by applying the proposed new SW decoding strategy, while preserving good analysis accuracy ($F_{score} \approx 0.9$) and frame error rate performance (FER $\leq 2 \times 10^{-4}$ for high SNRs $\geq$ 4dB). In future work, the proposed feature selection and coding strategy will be implemented and tested in real-time settings, using embedded camera sensors and Raspberry Pi boards, to prove its efficiency in terms of network lifetime maximization and decoding delay minimization for real-time applications.

REFERENCES

[1] H. Kavalionak, E. Carlini, A. Lulli, C. Gennaro, G. Amato, C. Meghini, and L. Ricci, "A prediction-based distributed tracking protocol for video surveillance," in *Proc. IEEE 14th Int. Conf. Netw., Sens. Control (ICNSC)*, May 2017, pp. 140–145.

[2] G. Leone, D. Moroni, G. Pieri, M. Petracca, O. Salvetti, A. Azzarà, and F. Marino, "An intelligent cooperative visual sensor network for urban mobility," *Sensors*, vol. 17, no. 11, p. 2588, Nov. 2017.
[3] N. Bendimerad and B. Kechar, "Rotational wireless video sensor networks with obstacle avoidance capability for improving disaster area coverage," *J. Inf. Process. Syst.*, vol. 11, no. 4, pp. 509–527, 2015.
[4] J. Zou, H. Xiong, C. Li, R. Zhang, and Z. He, "Lifetime and distortion optimization with joint source/channel rate adaptation and network coding-based error control in wireless video sensor networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1182–1194, Mar. 2011.
[5] O. Alaoui-Fdili, F.-X. Coudoux, Y. Fakhri, P. Corlay, and D. Aboutajdine, "Energy-efficient joint video encoding and transmission framework for WVSN," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4509–4541, Feb. 2018.
[6] H.-W. Kim and A. Kachroo, "Low power routing and channel allocation of wireless video sensor networks using wireless link utilization," *Adhoc Sensor Wireless Netw.*, vol. 30, pp. 83–112, Jan. 2016.
[7] N. Khernane, J.-F. Couchot, and A. Mostefaoui, "Routing impact on network lifetime maximization using power/rate trade-off in WVSN," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 97–102.
[8] N. Cen, Z. Guan, and T. Melodia, "Compressed sensing based low-power multi-view video coding and transmission in wireless multi-path multi-hop networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3122–3137, Sep. 2022.
[9] A. Bouchemel, D. Abed, and A. Moussaoui, "Enhancement of compressed image transmission in WMSNs using modified μ-nonlinear transformation," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 934–937, May 2018.
[10] S. M. Aziz and D. M. Pham, "Energy efficient image transmission in wireless multimedia sensor networks," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1084–1087, Jun. 2013.
[11] A. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Compress-then-analyze vs. analyze-then-compress: Two paradigms for image analysis in visual sensor networks," in *Proc. IEEE 15th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2013, pp. 278–282.
[12] L. Baroffio, J. Ascenso, M. Cesana, A. Redondi, and M. Tagliasacchi, "Coding binary local features extracted from video sequences," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2794–2798.
[13] N. Monteiro, C. Brites, F. Pereira, and J. Ascenso, "Multi-view distributed source coding of binary features for visual sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 2807–2811.
[14] G. Lin, C. Fan, H. Zhu, Y. Miu, and X. Kang, "Visual feature coding based on heterogeneous structure fusion for image classification," *Inf. Fusion*, vol. 36, pp. 275–283, Jul. 2017.
[15] L. Bondi, L. Baroffio, M. Cesana, A. Redondi, and M. Tagliasacchi, "Multi-view coding of local features in visual sensor networks," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun. 2015, pp. 1–6.
[16] D. Van Opdenbosch and E. Steinbach, "Collaborative visual SLAM using compressed feature exchange," *IEEE Robot. Autom. Lett.*, vol. 4, no. 1, pp. 57–64, Jan. 2019.
[17] A. E. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Multi-view coding and routing of local features in visual sensor networks," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2016, pp. 1–9.
[18] P. Monteiro, J. Ascenso, and F. Pereira, "Local feature selection for efficient binary descriptor coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4027–4031.
[19] N. Monteiro, C. Brites, F. Pereira, and J. Ascenso, "Multi-view distributed coding and selection of local binary features," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2016, pp. 1–6.
[20] C. M. Christoudias, R. Urtasun, and T. Darrell, "Unsupervised feature selection via distributed coding for multi-view object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[21] S. Eleuch, N. Khouja, S. Milani, T. Erseghe, and F. Tlili, "A study on the impact of multiview distributed feature coding on a multicamera vehicle tracking system at roundabouts," *IEEE Access*, vol. 10, pp. 39502–39517, 2022.
[22] N. Cen, Z. Guan, and T. Melodia, "Interview motion compensated joint decoding for compressively sampled multiview video streams," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1117–1126, Jun. 2017.
[23] S. Milani and G. Calvagno, "Distributed video coding based on lossy syndromes generated in hybrid pixel/transform domain," *Signal Process., Image Commun.*, vol. 28, no. 6, pp. 553–568, Jul. 2013.
[24] S. Milani, "A distributed source autoencoder of local visual descriptors for 3D reconstruction," *Pattern Recognit. Lett.*, vol. 146, pp. 193–199, Jun. 2021.
[25] S. Eleuch, N. Khouja, T. Erseghe, and F. Tlili, "Feature-based vehicle tracking at roundabouts in visual sensor networks," in *Proc. 17th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Jul. 2020, pp. 167–172.
[26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jan. 2008.
[27] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[28] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
[29] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[30] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
[31] M. Eldib, F. Deboeverie, D. V. Haerenborgh, W. Philips, and H. Aghajan, "Detection of visitors in elderly care using a low-resolution visual sensor network," in *Proc. 9th Int. Conf. Distrib. Smart Cameras*, Sep. 2015, pp. 56–61.
[32] P. Jin, P. Liu, and X. Cheng, "Safety for pedestrian recognition in sensor networks based on visual compressive sensing and adaptive prediction clustering," *Saf. Sci.*, vol. 117, pp. 10–14, Aug. 2019.
[33] J. Fernández-Berni, R. Carmona-Galán, J. F. Martínez-Carmona, and A. Rodríguez-Vázquez, "Early forest fire detection by vision-enabled wireless sensor networks," *Int. J. Wildland Fire*, vol. 21, no. 8, pp. 938–949, 2012.

SALMA ELEUCH received the degree in telecommunication engineering from the Higher School of Communication of Tunis (SUP'COM), Tunisia, in 2017. She is currently pursuing the joint Ph.D. degree with SUP'COM and the University of Padova, Italy. In 2016, she participated in the Erasmus Mobility Program with the University of Padova, where she carried out her master's thesis in wireless sensor networks. Her current research interests include visual sensor networks, smart traffic monitoring, distributed coding, and distributed optimization.



NADIA KHOUJA received the Diploma degree in communications engineering, the master's degree in communications, and the Ph.D. degree in communication sciences and information technologies from the High Communications School of Tunis, Tunisia, in 2002, 2006, and 2011, respectively. She worked as a Research and Development Engineer at STMicroelectrics, from 2002 to 2007, and worked on fields related to SoC and NoC modelization, and validation. She is currently an Associate Professor of telecommunications with the Institut Superieur des Etudes technologiques en communications de Tunis, ISETCOM, and a member of the GRESCOM Laboratory, SUP'COM, University of 7 November at Carthage, Tunisia. Her research interests include digital filtering, power consumption analysis, VLSI and embedded processors circuits, FEC decoding algorithms, and architectures.

SIMONE MILANI (Member, IEEE) received the Laurea degree in telecommunication engineering and the Ph.D. degree in electronics and telecommunication engineering from the University of Padova, Padova, Italy, in 2002 and 2007, respectively. He was a Visiting Ph.D. Student at the University of California at Berkeley, Berkeley, CA, USA, in 2006. He was a Consultant at STMicroelectronics, Agrate, Italy. He was a Postdoctoral Researcher at the University of Udine, Udine, Italy, the University of Padova, and the Politecnico di Milano, Milan, Italy, from 2007 to 2013. From 2013 to 2020, he was an Assistant Professor with the Department of Information Engineering, University of Padova, where he is an Associate Professor. His research interests include digital signal processing, image and video coding, 3-D video processing and compression, joint source-channel coding, robust video transmission, distributed source coding, multiple description coding, and multimedia forensics.

TOMASO ERSEGHE received the Laurea (M.Sc.) and Ph.D. degrees in telecommunication engineering from the University of Padova, Italy, in 1996 and 2002, respectively. From 1997 to 1999, he was at Snell and Wilcox—an English broadcast manufacturer. From 2003 to 2017, he was an Assistant Professor (Ricercatore) at the Department of Information Engineering, University of Padova, where he is an Associate Professor. His research interests include coding in the finite block-length regime, social network analysis and network science, distributed algorithms, smart grid optimization, ultra-wideband transmission systems design, spectral analysis of complex modulation formats, fractional Fourier transforms and their applications, image processing, and compression.

FETHI TLILI received the degree in electrical engineering, in 1990, the Ph.D. degree in electrical engineering, in 2000, and the University Habilitation degree in telecommunications, in 2011. He has been a Professor with the Higher School of Communications of Tunis (SUP'COM), Tunisia, since 1991, and a Researcher with the GRES'COM Laboratory, since 2012. He served as a senior consultant and the expert in CPL modems design, video technology, and radar processing for several international companies. His research interests include embedded systems, digital communications systems, HW architectures for signal processing, video technology, and radar processing for automotive systems.