Cooperative Features Extraction in Visual Sensor Networks: a Game-Theoretic Approach

Alessandro Enrico Redondi, Luca Baroffio, Matteo Cesana, Marco Tagliasacchi
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
{name.surname}@polimi.it

ABSTRACT

Visual Sensor Networks consist of several camera nodes that perform analysis tasks, such as object recognition. In many cases camera nodes have overlapping fields of view. Such overlap is typically leveraged in two different ways: (i) to improve the accuracy/quality of the visual analysis task by exploiting multi-view information or (ii) to reduce the consumed energy by applying temporal scheduling techniques among the multiple cameras. In this work, we propose a game theoretic framework based Nash Bargaining Solution to bridge the gap between the two aforementioned approaches. The key tenet of the proposed framework is for cameras to reduce the consumed energy in the analysis process by exploiting the redundancy in the reciprocal fields of view. Experimental results confirm that the proposed scheme is able to improve the network lifetime, with a negligible loss in terms of visual analysis accuracy.

Categories and Subject Descriptors

[**Computer systems organization**]: Embedded and cyber-physical systems—*Sensor networks*

Keywords

Visual Sensor Networks, Game Theory, Nash Bargaining Solution, Multi-view Object Recognition

1. INTRODUCTION

In recent years, several research efforts have flourished to enable classical Wireless Sensor Networks (WSNs) with vision capabilities, giving rise to the so-called Visual Sensor Networks (VSNs). VSNs bear the very same characteristics and constraints of wireless sensor networks, including the limitations in terms of transmission bandwidth, processing power and energy budget. Differently from classical WSNs, in VSNs some nodes are geared with cameras, to support advanced tasks such as surveillance and environmental monitoring [2]. Such apparently small difference induces novel challenges in the design process of VSNs, since acquisition, processing and transmission of multimedia flows are resource-eager operations which are at odds with the resource-constrained environment typical of WSNs.

The traditional system design for VSNs follows a *compress-then-analyze* (CTA) paradigm, where images (or videos) are acquired and compressed locally at the camera nodes, and then transmitted to one or multiple information sinks which perform the specific analysis tasks (video surveillance, face detection, object recognition, etc...). Recently, a paradigm shift has emerged: according to the *analyze-then-compress* (ATC) paradigm, the visual content is processed locally at the camera nodes, to extract a concise representation constituted by local visual features. In a nutshell, salient keypoints are detected in the acquired image. Then, for each keypoint a visual feature is computed by properly summarizing the photometric properties of the patch of pixels around the keypoint. Such features are then compressed and transmitted to the sink for further analysis. Since the features-based representation is usally more compact than the pixel-based one, the ATC approach is particularly attractive for those scenarios in which bandwidth is scarce, like VSNs [9].

Here we consider a reference scenario where a VSN is deployed to perform object recognition according to the ATC paradigm. In this scenario, each camera extracts visual features from the detected objects and transmits them to a central controller. There, the received features are matched with a database of labeled features from known objects to find the most similar one. We focus on the case of cameras with overlapping fields of view (FoVs): such a case may be encountered when cameras are densely deployed, as it happens in many surveillance applications [12].

Regardless of the specific application, the availability of multiple cameras capturing overlapping views of a scene is generally induced and/or exploited to improve the performance of the specific visual task. As an example, for the case of object recognition, multiple views of the same scene can provide obvious walk-arounds to occlusions. However, such performance improvement requires the cameras to be active (acquiring and processing) concurrently with additional costs in terms of network infrastructure and overall energy consumption.

This work analyses the accuracy/energy consumption trade-off involved in object recognition tasks performed by multiple cameras with partially overlapping FoVs. To this extent, we propose a game theoretic framework to model the cooperative visual feature extraction process, and we resort to the Nash Bargaining Solution (NBS) to steer the coopera-

ICDSC 2015 Sevilla, Spain

DOI: http://dx.doi.org/10.1145/2789116.2789124

tive processing. The key tenet of the proposed framework is to reduce the energy consumption for feature extraction by exploiting the redundancy in the reciprocal FoVs. The proposed scheme is then applied to different multi-view image datasets to assess its performance. Experimental results confirm that the proposed coordination scheme reduces the energy consumption with respect to the case in which multiple cameras process the whole input image, with a negligible loss in the achieved quality.

The paper is organized as follows: Section 2 overviews the related work; Section 3 provides some background notions on the reference object recognition pipeline and introduces the reference VSN topology. Section 4 describes the game theoretic-framework for cooperative object recognition in VSNs and Section 5 contains the performance evaluation of the proposed scheme in different network/dataset conditions. Concluding remarks are given in Section 6.

2. BACKGROUND AND RELATED WORK

In the last few years, an increasing number of works have faced the problem of managing VSNs featuring cameras with overlapping FOVs. The main focus has been given to coverage problems, which are critical in monitoring applications. In [1] the authors propose solutions to maximize the visual coverage with the minimum number of sensor, assuming to have cameras with tunable orientations. Wang and Gao in [12] propose a novel model called full-view coverage, observing that the viewing direction of a camera willing to recognize an object should be sufficiently close to the facing direction of that object. The same concept is leveraged in [13] where the authors present a method to select camera sensors from a random deployment to form a virtual barrier made of cameras for monitoring tasks. As a result, many redundant cameras (i.e., cameras with overlapping FoVs) might be selected. Since VSNs are battery operated, it is imperative to optimize their operation, thus maximizing their lifetime. In most of the works that deal with coverage, lifetime is defined as the amount of time during which the network can satisfy its coverage objective. With this definition, the approach traditionally used is to leverage the redundancy resulting from random deployment and organize redundant cameras in clusters. Then, coordination can be applied among cluster members, by putting to sleep some nodes while others sense the environment [3]. On the contrary, fusing information from multiple views of the same object, can improve the performance of visual analysis tasks. As an example, barrier coverage applications benefit from having multiple views from distinct active cameras. Similarly, Naikal et al. [7] propose a distributed object recognition system for VSNs where visual features extracted from multiple views of the same objects are leveraged to improve the efficiency of object recognition. Summarizing, there exists a dichotomy between the need of extending lifetime (which calls for de-activating camera nodes) and the need of improving the accuracy of the specific visual task (which requires many cameras to be active at the same time). To our knowledge, the available literature tends to focus on one of these two contrasting objectives; differently, we aim at gauging a more thorough analysis of the quality/lifetime tradeoff in VSNs by relying on a game theoretic framework. Game theory has recently been applied to VSNs for solving camera assignments problems [6] and for resource management optimization [8].

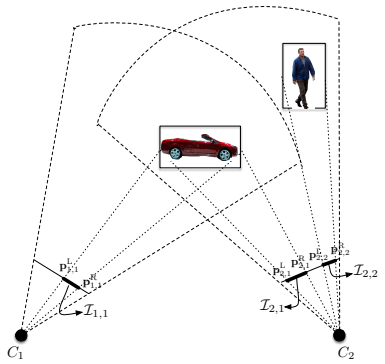


Figure 1: The two cameras C_1 and C_2 detect objects in their fields of view, and identify a BBx for each detected object. After exchanging the top-left and bottom-right pixel coordinates of each BBx, $\mathcal{I}_{1,1}$ and $\mathcal{I}_{2,1}$ are identified as corresponding image portions.

3. REFERENCE SCENARIO

We consider a scenario where N wireless camera nodes with fully or partially overlapping FoVs are able to communicate with each other, i.e., they are in direct transmission range. Figure 1 shows a descriptive example for $N = 2$: each camera acquires an image from the environment and has to perform local visual features extraction.

3.1 Non-cooperative visual sensor networks

In the case of non-coordinated networks, each camera independently performs the following steps:

- *Object detection:* the acquired image is pre-processed with foreground detection/background subtraction techniques to detect possible objects in the field of view. Typically, a bounding box (BBx) is drawn around each detected objects. Note that multiple objects may be detected in one image, as it is the case of camera C_2 in Figure 1. Therefore, we denote as $\mathcal{I}_{i,j}$ the portion of image contained in the j-th of the i-th camera, $i \in 1 \dots N$. Each BBx $\mathcal{I}_{i,j}$ is determined by its top-left and bottom right coordinates, namely $\mathbf{p}^{\mathrm{L}}_{i,j}$ and $\mathbf{p}^{\mathrm{R}}_{i,j}$, and the number of pixels contained in a BBx is denoted as $P_{i,j}$

- *Features extraction:* the pixels corresponding to each BBx are processed by means of a local features extraction algorithm. Such step encodes the photometric properties of each detected object in a representation which is: (i) generally more concise than the pixel-domain one and (ii) robust to several image transformations (scale, rotation, illumination changes) and thus ideal for being used for recognition tasks.

- *Features transmission:* the local features extracted from each BBx are transmitted to a remote controller, where they are matched against a database of labelled features and object recognition is performed.

Clearly, all the aforementioned steps require non negligible energy to be performed. Since in VSNs camera nodes are battery-operated, it is imperative to optimize the process of feature extraction and transmission to limit the corresponding energy consumption. In this work, we leverage the fact that cameras have overlapping FoVs to set up a

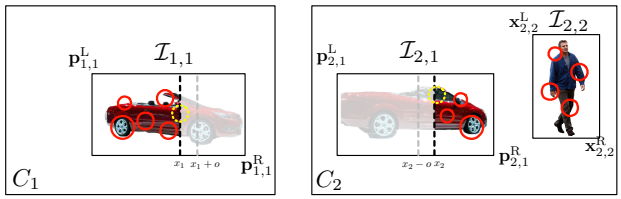


Figure 2: The BBx detected on the images acquired by C_1 and C_2, referring to the scenario in Fig. 1. Only for the corresponding BBx (the ones containing the car), an image sub-portion is selected from each camera for features extraction (red solid circles). Features near the splitting border (yellow dashed circles) are correctly detected, but can't be computed unless an offset region is added to the BBx.

cooperative framework that enables to save energy without sacrificing the visual analysis performance. In particular, we posit that features extracted from different views of the same object share a high degree of similarity. Therefore, different cameras may agree to share the features extraction task by processing only sub-portions of the detected object: since the energy needed to perform features extraction depends primarily on the number of processed pixels [4], such a cooperative approach is expected to provide notable energy savings with respect to the non-cooperative case, i.e., when each camera processes the entire bounding box.

3.2 Cooperative visual sensor networks

One main condition that must hold in order to enable the cooperative framework is that the cameras willing to cooperate are indeed looking at the same object. This check may be easily accomplished if the cameras are *calibrated*. In this case, the geometrical relationship between the two cameras is algebraically represented by the so-called fundamental matrix $\mathbf{F}$, available to both cameras and allowing to check if a point $\mathbf{p}$ (in pixel coordinates) in the first view corresponds to a point $\mathbf{p}'$ in the second view (see Figure 1), through the well known fundamental matrix equation:

$$\mathbf{p}'^{\mathbf{T}}\mathbf{F}\mathbf{p} = 0. \tag{1}$$

Therefore we propose the following additional steps to set up the cooperative framework:

- *Geometric consistency:* After object detection, calibrated cameras exchange the top-left and bottom-right pixel coordinates of the detected BBx in their respective fields of view. Then, they use equation (1) to check if they are looking at the same object. Note that this process is not limited to pairs of cameras, but may be easily extended to networks of multiple camera nodes with overlapping FoVs by relying on transitivity.
- *Bounding box splitting:* having identified a common object in their FoVs, the cameras may select a sub-portion of their own BBx instead of processing them entirely. As illustrated in Figure 2, for the case of two cameras, we assume that the leftmost camera selects from the leftmost region of its BBx up to x_1, while the rightmost camera processes from x_2 to the right end, where x_1 and x_2 can be expressed as proportions of the BBx area that are processed by the first and the second camera, respectively.. That is, x_i is in the range from 0 (when the i-th camera does not perform any processing) to 1 (when the i-th camera processes its entire BBx). Note that, without loss of generality, splitting may be applied in the vertical direction as well. In the following section, we propose a game theoretic approach for determining the proportion of the BBx to be processed on each camera, i.e., the values of x_i, $i = 1 \dots N$. Once such values are computed, each camera may extract features from the reduced BBx and transmit them to the central controller. A reasonable constraint for the variables x_i is that they sum up to 1, i.e., visual features are extracted from the entire object, although in different views. However, it is important to note that image splitting may negatively affect the features extraction process. As illustrated in Figure 2, this is due to the fact that the extraction of one visual feature requires the processing of a patch of pixels around the corresponding keypoint. If the keypoint is detected close to the splitting line, there may not be enough pixels to perform the feature extraction. In the case of a very discriminative feature being close to the splitting line, such approach may negatively affect the performance of object recognition. To overcome this issue, an offset is added to the variables x_i. In the following, we denote by o such required offset, normalized with respect to the total image size.

4. GAME-THEORETIC MODELS

The reference scenario described in Section 3 can be modeled as a game among N cameras which have to decide the portion of the common bounding box they need to process. Let $\mathbf{x} = (x_1, x_2, \dots x_N)$ be an outcome of the game, being x_i the portion of the BBx which is assigned for processing to camera i, with $x_i \geq 0$ and $\sum_{i=1}^{N} x_i = 1$. Let $\mathcal{X}$ be the set of all possible outcomes of the game. Let us further define an utility function $u_i(\mathbf{x})$ which represents the *preference* for camera i on the outcome $\mathbf{x}$. The set of possible payoff vectors is defined as $\mathcal{U} = \{u_1(\mathbf{x}), u_2(\mathbf{x}), \dots, u_N(\mathbf{x})\}$.

In the reference scenario, it is reasonable to bind the utility function $u_i(\mathbf{x})$ to the energy consumed in the feature extraction process. The energy consumed for extracting features increases linearly with the number of processed pixels [4]. Thus, we model the energy consumption as a linear function of the number of pixel processed defined as:

$$E_i(x_i) = P_i(a_i x_i + b_i), \tag{2}$$

where the parameters a_i and b_i depend on the particular processor available on the i-th camera, and P_i is the size in pixels of the bounding box currently under processing. We define the utility function for camera i as:

$$u_i(\mathbf{x}) = E_i\Big(\sum_{k=1, k\neq i}^{N} x_k\Big) \tag{3}$$

Intuitively, the utility function for the i-th camera is the amount of energy that camera i saves through cooperative processing. The scenario under consideration can be modeled as a bargaining problem with two main ingredients:

- **Feasibility set:** the convex set $\mathcal{F} \subseteq \mathbb{R}^N$ including all the possible payoff vectors $\mathbf{u} = (u_1, u_2, \dots, u_N)$ defined by Eq. (3);

- **Disagreement point:** the value of the utility function the players are expected to receive if the negotiation breaks down $\mathbf{d} = (d_1, \ldots, d_N)$; in our case, if negotiation breaks down, each camera has to process the entire BBx, thus its utility (e.g., spared energy) at the disagreement point is null, $d_i = 0$, $i = 1 \ldots N$.

The Feature Extraction Bargaining Problem (FEBP) can be defined as the tuple $(\mathcal{F}, \mathbf{d})$. A solution concept which can be applied to such game theoretic scenario is the generalized Nash Bargaining Solution (NBS) which provides an axiomatic solution to the bargaining, further providing an operative method to derive it. Formally, the NBS defines an agreement point (bargaining outcome) $\mathbf{x}_{\text{NBS}}$ which verifies the following four axioms:

1. **Rationality:** $u_i(\mathbf{x}_{\text{NBS}}) > d_i$, $i = 1 \ldots N$, i.e., no player would accept a payoff that is lower than the one guaranteed to him under disagreement;
2. **Pareto optimality:** under the optimality conditions, the payoff of each player cannot be further improved without hurting other players' ones;
3. **Symmetry:** if the players are undistinguishable, the agreement should not discriminate between them;
4. **Independence of irrelevant alternatives:** the solution of a bargaining problem does not change as the set of feasible outcomes is reduced, as long as the disagreement point remains the same, and the original solution feasible.

Since $\mathcal{X}$ is compact and convex and the utility functions $u_i(\mathcal{X})$ are concave and upper bounded, the generalised NBS for the bargaining problem is the unique solution of the following optimization problem [11]:

$$\begin{aligned} &\text{maximize} \prod_{i=1}^{N} (u_i(x_i) - d_i)^{\alpha_i} \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0 \quad \forall i \end{aligned} \tag{4}$$

The exponents α_i represent the *bargaining power* of each camera, and are chosen such that $\sum_{i=1}^{N} \alpha_i = 1$. A natural choice for the bargaining powers α_i is to relate them to the residual energy of each camera $\bar{E}_i$. In particular, a desirable condition is that a camera bargaining power increases as its residual energy decreases (i.e., cameras close to deplete their energy are eagerer to cooperate). Thus, we define the bargaining powers as:

$$\alpha_i = \frac{\bar{E}_i^{-1}}{\sum_{i=1}^{N} \bar{E}_i^{-1}} \tag{5}$$

5. PERFORMANCE EVALUATION

We are interested in assessing the performance of the cooperative framework in terms of object recognition accuracy and energy efficiency. To this extent, we have implemented the full pipeline of a typical object recognition task based on BRISK [5] visual features: camera nodes acquire a *query* image, extract visual features from it, and transmit the features to a sink node where object recognition is performed. There, the received features are matched against features extracted from a database of images. Matching consists in pair-wise comparisons of features extracted from, respectively, the query and database image. The Hamming distance is adopted to measure the similarity between BRISK visual features extracted from the image and the one contained in the database. Two features are labeled as matching if their distance is below a pre-defined threshold. Additionally, a geometric consistency check step based on RANSAC is applied to filter out outliers. Hence, the images in the database can be ranked according to the number of matches with the query image.

5.1 Accuracy Evaluation

Average Precision (AP) is commonly adopted to assess the performance of object recognition/image retrieval. Given a query q, AP is defined as:

$$AP_q = \frac{\sum_{k=1}^{n} P_q(k) r_q(k)}{R_q}, \tag{6}$$

where $P_q(k)$ is the precision (i.e., the fraction of relevant documents retrieved) considering the top-k results in the ranked list; $r_q(k)$ is an indicator function which is equal to 1 if the item at rank k is relevant for the query, and zero otherwise; R_q is the total number of relevant documents for the query q and n is the total number of documents in the list. The Mean Average Precision (MAP) for a set of Q queries is the arithmetic mean of the APs across different queries:

$$MAP = \frac{\sum_{q=1}^{Q} AP_q}{Q} \tag{7}$$

5.2 Energy Evaluation

Energy efficiency is captured by estimating the lifetime L of the system, that is the number of consecutive queries (images) which can be processed until one of the camera nodes depletes its energy. That is:

$$L = \min \frac{E_i^{\text{budget}}}{\frac{1}{Q} \sum_{q=1}^{Q} E_{i,q}}, \tag{8}$$

where E_i^{budget} is the energy budget of the i-th camera and $E_{i,q}$ is the energy required for processing the q-th query on the i-th camera. To characterize the per-query energy consumption of a camera, we rely on the following energy model:

$$E_{i,q} = E^{\text{acq}} + P^{\text{cpu}} \left[t_{i,q}^{\text{bb}} + t_{i,q}^{\text{det}} + t_{i,q}^{\text{desc}} \right] + E^{\text{tx}} r M_i, \tag{9}$$

being E^{acq} the energy required for acquiring one image, P^{cpu} the power consumption of the CPU of each camera and $t_{i,q}^{\text{bb}}, t_{i,q}^{\text{det}}$ and $t_{i,q}^{\text{desc}}$ the times taken by the i-th camera to identify the bounding boxes, detect keypoints and extract features for the q-th query, respectively. The energy cost of transmitting the extracted features is captured by the last term of (9), where E^{tx} is the energy cost of transmitting one bit, r is the dimension in bit of each visual feature, and M_i is the number of features detected by the i-th camera. The values used for the energy costs are based on a Visual Sensor Node platform based on a BeagleBone linux computer [10] and are reported in Table 1.

Table 1: Parameters used for the energy evaluation

Name	Symbol	Value
CPU power	P^{cpu}	1.75 W
Energy budget	E^{budget}	20 KJ
Acquisition cost	E^{acq}	10^{-3} J/frame
Transmission cost	E^{tx}	2.2×10^{-7} J/bit
Feature size	r	512 bit

5.3 Experimental Methodology

The evaluation has been carried out on several VSN topologies, each one consisting of a pair of camera nodes characterized by a different geometrical relationship, and by relying on different image datasets. From each dataset, we selected one common set of images as the reference database for the object recognition task and several set of images as query datasets. The query datasets are selected so as to mimic different camera geometries:

- *COIL100*[1]: this image database contains 100 objects, each captured at 72 different poses. Each pose of an object is obtained by rotating the object by 5 degrees. For each object, the reference database contains three images corresponding to the views at 0° and $\pm10^\circ$. Five different camera geometries are tested as query datasets, taking for each object the couple of images at $\pm5^\circ$, $\pm15^\circ$, $\pm20^\circ$,$\pm25^\circ$ and $\pm30^\circ$. We refer to such experiments with the label COIL-X, where X is in the set {5,15,20,25,30}.

- *ALOI*[2]: an image collection of one-thousand small objects. Similarly to the COIL-100 dataset, each object is captured at 72 different poses obtained by rotating the object by 5 degrees each time. The reference database and the test sets are obtained in the same way as for the COIL-100 dataset (i.e., five different camera configurations, each one with an increasing rotation). Again, we refer to such experiments with the label ALOI-X, where X is in the set {5,15,20,25,30}.

- *ANT66*[3]: A novel image database containing 66 objects, each one captured by two camera pairs with overlapping FoVs. The reference database contains one image per object and two camera geometries are available, which we refer to as ANT-0 and ANT-15. In ANT-0, the inter-camera geometry is a pure translation, while in ANT-15 the two cameras are translated and rotated by $+15^\circ$ and -15° degrees with respect to the object's main axis.

For each one of the twelve different camera topologies, the NBS-based cooperative framework is evaluated as follows:

1. For each of the two cameras, load a query image q from the current test set.

2. Find the image splitting x_1 and x_2 by solving the features extraction bargaining problem through the generalized NBS according to equation (4).

3. Extract BRISK features from the sub-portions defined by $x_1 + o$ and $x_2 - o$. Compute the per-camera energy consumption E_1 and E_2 as in (9), and the average precision AP_q as in (6). To compute the AP, the feature sets from the two camera views are independently matched against the reference dataset, and geometrically verified through RANSAC. The number of matches for the q-th query couple is then computed by summing the matches from the two independent views.

4. Update residual energies for the two cameras, $\bar{E}_1$ and $\bar{E}_2$.

5. Repeat steps 1-4 until (i) one of the two cameras deplete its energy or (ii) all queries in the datasets have been processed. Compute the MAP as in (7) and the system lifetime as in (8).

The entire process is repeated for increasing values of the offset o. For each camera topology, we also compared the NBS-based cooperative framework against the following two baseline scenarios:

- *Temporal Scheduling (TS):* at each query, only one camera among the ones with overlapping FoVs acquires the image and performs object recognition [3]. We select which camera should operate according to the maximum residual energy.

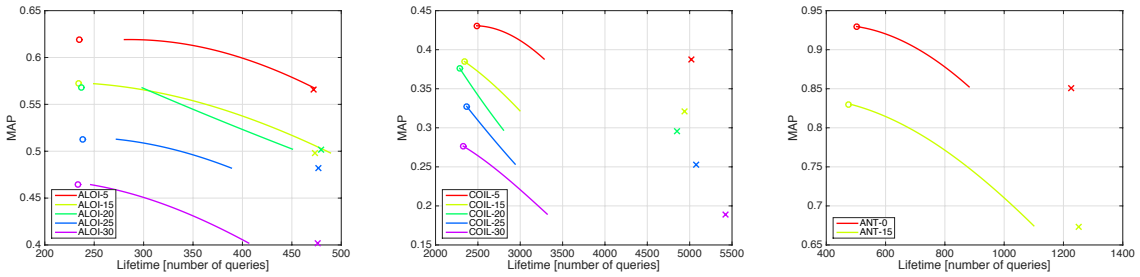
- *Multi View object recognition (MV):* following the approach in [7], at each query all cameras acquire an image, extract the corresponding features and transmit them to the sink node. Similarly to the NBS, features matching is performed by using the features extracted from the two camera views.

5.4 Experimental Results

Figure 3 illustrates the energy/accuracy trade-offs obtained by running the aforementioned experiments on the three reference datasets. Each figure reports: (i) the performance of the NBS cooperative framework for different values of the overlap o (solid line), (ii) the performance of the temporal scheduling approach (colored cross) and (iii) the performance of the multi-view approach (colored circle), for the different camera topologies.

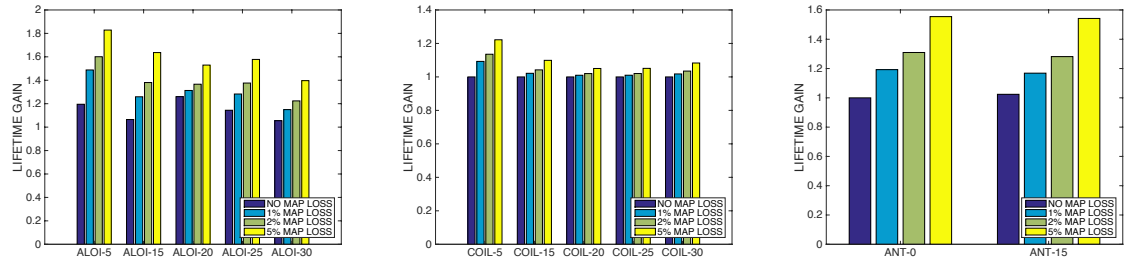
Expectedly, using the MV approach leads to higher MAP than applying temporal scheduling at the expense of a notably higher overall energy consumption in all the considered topologies. Referring to Figure 3(a), it can be observed that the NBS-based solution allows to efficiently trade-off MAP for energy, that is, it allows to reduce the consumed energy by a factor of 2 approximately (with respect to the MV approach) with a limited MAP loss (around 5%). As the reference topologies become more challenging, that is, with increasing angle of displacement between the two cameras, the trade off between the MAP and system lifetime becomes less favorable, and the energy savings diminishes for a given MAP loss. Figure 3(b) and 3(c) report the same analysis for the COIL100 and the ANT66 datasets, respectively. Figure 4 summarizes the numerical analysis by reporting, for all the considered data sets, the gain in the system lifetime for a given tolerable MAP loss, always with respect to the MV scenario (two cameras always active). For a given dataset configuration, the lifetime gain increases as the tolerated level of MAP loss also increases. In general, for a given tolerated MAP loss the achievable lifetime gain tends to decrease as the camera geometry becomes more challenging.

[1] http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[2] http://aloi.science.uva.nl/
[3] http://www.greeneyesproject.eu/



(a) ALOI (b) COIL100 (c) ANT66

Figure 3: Accuracy (MAP) vs energy (Lifetime) trade-off for the TS approach (cross), the MV approach (circle) and the NBS (solid line). Different camera geometries are illustrated with different colors.



(a) ALOI (b) COIL100 (c) ANT66

Figure 4: NBS lifetime gain with respect to the MV approach for different values of tolerated MAP loss.

6. CONCLUSIONS

We have presented a cooperative framework for object recognition through features extraction in VSNs, when cameras have overlapping FoVs. By relying on the generalized NBS, we were able to trade off system lifetime for task accuracy, improving the network lifetime with a negligible loss in the achieved visual analysis accuracy. Future works include the evaluation of the proposed method when using $N > 2$ cameras and for analysis tasks beyond object recognition.

7. ACKNOWLEDGEMENTS

The project GreenEyes acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number:296676.

8. REFERENCES

[1] J. Ai and A. A. Abouzeid. Coverage by directional sensors in randomly deployed wireless sensor networks. *Journal of Combinatorial Optimization*, 11:21–41, 2006.

[2] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury. A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4):921–960, 2007.

[3] M. Alaei and J. M. Barcelo-Ordinas. Node clustering based on overlapping fovs for wireless multimedia sensor networks. In *WCNC*, pages 1–6. IEEE, 2010.

[4] A. Canclini, M. Cesana, A. Redondi, M. Tagliasacchi, J. Ascenso, and R. Cilla. Evaluation of low-complexity visual feature detectors and descriptors. In *Digital Signal Processing (DSP), 2013 18th Intl. Conf. on*, pages 1–7, July 2013.

[5] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Proceedings of the 2011 Intl. Conf. on Computer Vision*, pages 2548–2555, 2011.

[6] Y. Li and B. Bhanu. Utility-based camera assignment in a video network: A game theoretic framework. *Sensors Journal, IEEE*, 11(3):676–687, March 2011.

[7] N. Naikal, A. Yang, and S. Sastry. Towards an efficient distributed object recognition system in wireless smart camera networks. In *Information Fusion (FUSION), 13th Conf. on*, pages 1–8, July 2010.

[8] K. Pandremmenou, L. P. Kondi, K. E. Parsopoulos, and E. S. Bentley. Game-theoretic solutions through intelligent optimization for efficient resource management in wireless visual sensor networks. *Signal Processing: Image Communication*, 29(4):472 – 493, 2014.

[9] 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 *IEEE Multimedia Signal Processing (MMSP), 2013*, pages 278–282, Sept 2013.

[10] A. Redondi, D. Buranapanichkit, M. Cesana, M. Tagliasacchi, and Y. Andreopoulos. Energy consumption of visual sensor networks: Impact of spatio-temporal coverage. *Circuits and Systems for Video Technology, IEEE Transactions on*, 24(12):2117–2131, Dec 2014.

[11] C. Touati, E. Altman, and J. Galtier. Generalized nash bargaining solution for bandwidth allocation. *Computer Networks*, 50(17):3242 – 3263, 2006.

[12] Y. Wang and G. Cao. On full-view coverage in camera sensor networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 1781 –1789.

[13] Y. Wang and G. Cao. Barrier coverage in camera sensor networks. In *Proceedings of the Twelfth ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing*, pages 12:1–12:10, 2011.