GREENEYES: ENABLING VISUAL ANALYSIS ON RESOURCE-CONSTRAINED WIRELESS SENSOR NETWORKS

L. Baroffio, L. Bondi, M. Cesana, A. Redondi, M. Tagliasacchi
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano - Italy

ABSTRACT

This demo showcases some of the results obtained by the GreenEyes project, whose main objective is to enable visual analysis on resource-constrained wireless sensor networks. The demo features a multi-hop visual sensor network operated by BeagleBones Linux computers with IEEE 802.15.4/802.11 communication capabilities, and capable of recognizing and tracking objects according to two different visual paradigms. In the traditional compress-then-analyze (CTA) paradigm, JPEG compressed images are transmitted through the network from a camera node to a central controller, where the analysis takes place. In the alternative analyze-then-compress (ATC) paradigm, the camera node extracts and compresses local binary visual features from the acquired images (either locally or in a distributed fashion) and transmits them to the central controller, where they are used to perform two demonstrative applications: object recognition/tracking and parking lot monitoring. We show that, in a bandwidth constrained scenario, the ATC paradigm allows to reach better results in terms of application frame rates and consumed energy, still ensuring excellent analysis performance.

***Index Terms*—** Binary Local Visual Features, Visual Sensor Networks, ARM, Object Recognition, Object Tracking, Parking Lot Monitoring

1. INTRODUCTION

Visual Sensor Networks (VSNs) have attracted the interest of researchers worldwide in the last few years, and are expected to play a major role in the evolution of the Internet-of-Things (IoT) paradigm with applications such as video surveillance, object and face recognition, object tracking and many others. Such visual tasks are typically accomplished through the extraction and analysis of global and local features from the pixel domain content: thus, they can be implemented in different ways in the VSN, depending on *where* in the network the task of feature extraction is performed.

In the traditional *compress-then-analyze* (CTA) approach, camera nodes acquire images and compress them relying on some sort of pixel level compression schemes (e.g., JPEG). The compressed images are then then delivered through the wireless sensor network to a central controller that extracts the features and performs visual analysis. As a consequence, depending on the amount of compression, the accuracy of the final analysis task might be significantly impaired. Moreover, when only the result of the visual analysis matters, transmitting image or video data retaining a pixel-level representation is inefficient in terms of computational and network resources used. For these reasons, the GreenEyes project considers an alternative approach in which, rather than transmitting pixel-level information, the visual content is processed by the sensing node to extract a compact representation, that is then transmitted for further analysis.. We call this approach *analyze-then-compress* (ATC). In this approach, image features are extracted by visual sensor nodes, encoded, and then delivered to the final destination(s) in order to enable higher level visual analysis tasks.

In this demo we showcase an efficient implementation of the ATC paradigm on a real visual sensor network, and we demonstrate its benefits compared to the traditional CTA paradigm in a bandwidth-limited scenario. We also show several key results of the GreenEyes project such as binary features encoding and distributed features extraction among neighboring nodes.

2. TECHNICAL DESCRIPTION

We demonstrate several novel solutions proposed by the GreenEyes project:

1. **Energy-efficient features extraction:** the complexity of the feature extraction algorithms at the base of ATC may be critical. In this demo we leverage a recent work of ours aimed at optimizing the BRISK features extraction algorithm [1] for ARM-based platforms. The BRISKOLA (BRISK Optimized for Low-Power ARM Architecture)[2] features extractor allows to obtain average speed-ups of 1.5 with respect to the original BRISK implementation.

2. **Lossless coding of binary features:** we also showcase a lossless entropy-coding scheme for compressing the extracted BRISKOLA features [3], which achieves bitrate reductions up to 20%.

3. **Distributed and cooperative features extraction:** camera nodes may also leverage the presence of neighboring network nodes to optimize the task of feature extraction. If neighbouring nodes are not equipped with a camera, they can still be exploited to reduce the overall processing time through offloading, following a Distributed Analyze-then-Compress (DATC) paradigm [4]. If they have a camera, and its field of view overlaps with the one of another camera node, cooperative features extraction schemes may be implemented.

3. IMPLEMENTATION

The demonstration is built on the following equipment:

- **Visual sensor nodes:** several battery-operated 720MHz ARM BeagleBone Linux computers which are geared with a Logitech USB camera to capture still images; the visual sensor nodes are also equipped with IEEE 802.15.4-compliant sensor nodes (TelosB platform or similar) to remotely transfer the visual content through low-power wireless links. The camera nodes are able to operate following the CTA, ATC and DATC paradigms. The choice of which paradigm to use is remotely controlled.

- **Cooperator nodes:** several battery-operated BeagleBone Linux computers similar to the visual sensor nodes but without sight capabilities. This type of nodes is used to implement the DATC paradigm.

- **Network infrastructure:** a network of battery-operated IEEE 802.15.4-compliant TelosB sensor nodes which is used to route the visual information to a central controller.

- **Central controller:** a laptop with IEEE 802.15.4 communication capabilities to receive the multimedia content transferred by the visual sensor nodes and to perform different visual analysis tasks: object recognition/tracking and parking lot monitoring. The central controller implements a graphical user interface which provides a highly interactive remote controller of the visual sensor nodes. The user can switch on the fly between the operating paradigms (CTA, ATC or DATC) and for each paradigm different parameters may be changed (e.g., JPEG quality factor in CTA mode, or the feature detection threshold and the maximum number of features to be transmitted in ATC mode). Moreover, the demonstrator estimates and displays: (i) the current frame rate, i.e., the maximum number of images which can be processed per unit time, under the different paradigms, (ii) the energy consumed per frame and (iii) the result of the visual analysis task. Such information is useful to assess the pros and cons of the different operational paradigms.

Application scenario: For the object recognition and tracking case, the demonstration scenario is composed of a LEGO model of a city containing several different objects. One of the camera node is mounted on a toy car which is able to move freely inside the city, in order to recognize and track the different objects. For the parking lot monitoring scenario, images from a parking lot are loaded into the testbed. A classification algorithm is then used to count the number of unoccupied parking spaces, either in CTA or ATC mode.

4. CONCLUSIONS AND FUTURE DEVELOPMENTS

The proposed demo showcases that, in the context of VSNs characterized by a limited transmission bandwidth, the ATC paradigm outperforms the traditional CTA paradigm in terms of the frame rate and consumed energy, while achieving the same visual analysis accuracy. Moreover, leveraging the presence of neighboring nodes for distributing the task of features extraction may lead to notable performance improvements. Future work will focus on other aspects related to multiple cameras in the VSN, such as inter-view features encoding. We also plan to extend the comparison between CTA and ATC to the case where temporal correlation between acquired images is exploited. A detailed video describing the demonstrator is available at `www.greeneyesproject.eu`

5. REFERENCES

[1] S. Leutenegger, M. Chli, and R. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” in *International Conference on Computer Vision*, 2011, pp. 2548–2555.

[2] L. Baroffio, A. Canclini, M. Cesana, A. Redondi, and M. Tagliasacchi, “Briskola: Brisk optimized for low-power arm architectures,” in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 5691–5695.

[3] A. Redondi, L. Baroffio, J. Ascenso, M. Cesana, and M. Tagliasacchi, “Rate-accuracy optimization of binary descriptors,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 2910–2914.

[4] Alessandro Redondi, Matteo Cesana, Marco Tagliasacchi, Ilario Filippini, György Dán, and Viktoria Fodor, “Cooperative image analysis in visual sensor networks,” *Ad Hoc Networks*, vol. 28, pp. 38–51, 2015.