
A. Canclini, L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi
Dipartimento di Bioingegneria, Elettronica e Informazione, Politecnico di Milano
P.za Leonardo da Vinci 32, Milano, Italy
lastname@elet.polimi.it

Abstract—This demo features a multi-hop Visual Sensor Network, capable of recognizing objects using two different visual paradigms. In the 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. Conversely, in the analyze-then-compress (ATC) paradigm, the camera node extracts and compress local visual features from the acquired images and transmits them to the central controller, where they are used to perform object recognition. We show that, in a bandwidth constrained scenario, the latter paradigm allows to reach higher application frame-rates, still ensuring excellent recognition results.

I. INTRODUCTION

Visual Sensor Networks (VSNs), composed of several inexpensive wireless camera nodes, are used to acquire images or video from the environment, which may be further processed to perform a broad range of complex visual analysis tasks, such as object recognition, event detection, localization and tracking, etc. The acquired visual information is typically delivered to a remote destination through low-power wireless transmissions (e.g., IEEE 802.15.4), and multiple visual or generic sensor nodes may be involved in the process of routing this information to its final destination. Due to their flexibility and low-cost, 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. We focus on a fundamental application of Visual Sensor Networks: the detection and recognition of objects. This is typically accomplished by extracting global or local distinctive features from the acquired images, and by matching these features against a database of pre-stored and labelled features to finally perform object recognition. Such visual task can be practically implemented in different ways in the VSN depending on where in the network the task of feature extraction is performed. Broadly speaking, two distinct paradigms are possible: compress-then-analyze (CTA) and analyze-then-compress (ATC) [1].

Traditionally, the CTA paradigm is adopted (Figure 1(a)). The acquired images are locally compressed at the camera sensor (e.g., using standard techniques like JPEG) and delivered through the wireless sensor network to a central controller which performs object recognition. The bitstream flowing in the network thus includes the compressed version of a pixel-representation of the acquired image. Object recognition performed at the central controller is thus based on a compressed and lossy representation of the original image, which might significantly impair the recognition accuracy. Also, several works in the past demonstrated that multi-hop image transmission in VSNs results in high latency and low application frame rates, due to the struggling between bandwidth availability and requirements [2]. Alternatively, in ATC (Figure 1(b)) image features are extracted locally at the multimedia sensor, compressed, and then delivered to the final destination in order to enable higher level visual analysis tasks. The key tenet is that most visual analysis tasks can be carried out based on a succinct representation of the image, which entails both global and local features, while it disregards the underlying pixel-level representation. This approach allows to perform object recognition at higher frame rates with respect to the CTA case, as the visual information transmitted through the network is much more compact.

![Diagram](image-url)

(a) Compress-then-analyze (CTA) paradigm

(b) Analyze-then-compress (ATC) paradigm

Fig. 1. The two different approaches to perform object recognition in visual sensor networks
II. Description of the Demonstrator

With reference to Figure 2, the demo testbed is composed of the following equipment:

- **Visual sensor node**: a battery-operated 720MHz ARM BeagleBone Linux computer which is geared with a Logitech USB camera to capture still images; the visual sensor node is also attached to a IEEE 802.15.4-compliant sensor node (TelosB platform or similar) to remotely transfer the visual content through low-power wireless links.

- **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 node and to perform visual analysis.

The visual sensor node is able to run both visual paradigms. Namely, as far as the CTA paradigm is concerned, the visual sensor node implements JPEG compression on the acquired image at different quality factors, which allows to trade off bandwidth with image quality. In the ATC case, the Binary Robust Invariant Scalable Keypoints (BRISK [3]) algorithm is implemented and used to extract local visual features. The camera node has also the ability to compress BRISK local features, by using a suitable encoding scheme to minimize the amount of bits required by each local feature without affecting the recognition capabilities [4]. At the central controller, the received features, in the ATC case, or the features extracted from the received JPEG image, in the CTA case, are matched against a database of labeled features, so that object recognition can be performed. The central controller also implements a graphical user interface which allows to visualize the result of object recognition and remotely control the operation of the visual sensor node, as illustrated in Figure 3; in details, the management console allows to change on the fly the type of paradigm, as well as all the most critical parameters of the two paradigms (quality factor for CTA, the number of features and the encoding scheme for the ATC). Moreover, in the ATC case, we implemented an engine to reconstruct a pixel-domain image from the knowledge of the local features only [5].

The demo experiment will showcase first a classical object recognition applications with the testbed being able to recognize the type of object which is seen by the visual sensor node. Then, the testbed will be also used to assess the maximum frame rate, that is, the maximum number of images which can be processed per unit time, under the two paradigms, showing that ATC outperforms CTA.

III. Demo Video

A detailed video of the demonstration is available at www.greeneyesproject.eu.

REFERENCES


