A Survey on Object Detection Methods in Visual Sensor Networks

Shamim Yousefi and Samad Najjar Ghabel

Faculty of Electrical and Computer Engineering, University of Tabriz, 29th Bahman Boulevard, Tabriz, Iran, Islamic Republic of

Abstract

Object detection is one of the major challenges in visual sensor networks (VSNs) which is set up in the monitoring applications. Many approaches proposed to solve the object detection problem in VSNs, considering diverse metrics such as reliability, energy consumption, detection accuracy and being real-time. In this paper, a survey on the object detection methods in visual sensor networks is presented for the first time. Furthermore, this paper classified the methods precisely. Two main object detection categories in VSNs that explored in this paper are conventional object detection methods and object detection approaches with the camera nodes involvement. To be more precise, presented survey promotes an overview of recent object detection methods' literature with their performance evaluation. Also, this research is challenging and the object detection issue in the visual sensor networks is open caused by differences in estimations and performance metrics. Therefore, the survey concludes with open research challenges.

Keywords

Detection Accuracy, Energy Consumption, Object Detection, Object Recognition, Visual Sensor Networks

1. Introduction

Recent advances in low-power and resource constraining self-organizing sensor nodes have led to the development of visual sensor networks (VSNs) [1, 2, 3]. The camera nodes that constitute the visual sensor networks are able to capture the multimedia data from the monitoring area in the form of conventional or infrared images and video streaming. Therefore, VSNs can lead to the development of monitoring and surveillance applications such as traffic control systems [4, 5], person locator services [6], industrial process control systems, seismic sensing and hazardous environment exploration [7] automated assistance for the elderly and family monitoring, biomedical health monitoring [8, 9] and virtual reality [10].

On most of the expressed applications, object detection and recognition are one of the major challenging issues in the visual sensor networks [11]. Two principal object detection categories in visual sensor networks that explored in this paper are conventional object detection methods and object detection approaches with the camera nodes involvement. The conventional object detection approaches in VSNs used various types of the background subtraction techniques [12] to recognize the difference between the background and foreground images. In such method, if the background subtraction result exceeds a pre-determined threshold, the camera node will detect moving objects within the monitoring area and sends the captured/difference image to the base station for recognition operations [13, 14]. In visual sensor networks, the data communication cost is usually much higher than the image processing cost [15]. Therefore, the

DOI:10.5121/ijassn.2016.6201
conventional object detection approaches are not suitable for monitoring and surveillance applications.

For reducing data transmission cost, some works involves the camera nodes to perform preprocessing after background subtraction and transmit only the bounding box of the objects to the network [16]. In these methods, if there are some non-object pixels between the objects in the foreground image, they are sent to the base station, too. As a result, this type of methods will be suitable on applications for detecting a single object. Some other works involve the camera nodes to perform preprocessing tasks, after background subtraction and send only the features or key-points of the objects to the base station [17, 18, 19]. The overall taxonomy of the object detection methods in visual sensor networks are shown in Figure 1.

![Figure 1: Overall Taxonomy of the Object Detection Methods in VSNs](image)

In the recent papers, various approaches for object detection in visual sensor networks are presented, but there is a deficiency of well-defined grouping of them based on applications' requirements. So, this survey presents taxonomy of object detection methods and discusses each approach under the appropriate category. The rest of the paper is organized as following: Section 2 provides a more detailed survey about the object detection methods in visual sensor networks and categorizes the approaches. In Section 3, reviewed methods are compared and the paper is concluded based on VSNs' applications.

2. **Taxonomy of Object Detection Approaches in VSNs**

Two main object detection categories in visual sensor networks that are examined in this section are conventional object detection methods and object detection methods with the camera nodes involvement. The conventional object detection approaches in VSNs used various types of background subtraction techniques [12] to recognize the moving objects in the camera node's field-of-view. In such methods, camera nodes send the captured/difference image that include objects to the base station [13, 14]. While object detection methods with the camera nodes involvement perform preprocessing tasks after background subtraction and send only the bounding box of the objects or the useful information of them into the network [16, 17, 18, 19].

2.1. **Conventional Object Detection Methods**

In most of the traditional visual sensor networks' applications, background subtraction techniques [12] are one of the common approaches to detect the presence of moving objects. These approaches are based on the difference between the background and foreground images [20]. At first, each camera node captures a reference image from its field-of-view without any moving objects, which called background image [21]. After saving the background image, the camera nodes periodically capture the foreground image [22] from its field-of-view.
In background subtraction approaches for object detection in VSNs, each camera node subtracts the background and foreground images. If the difference between the background and foreground images exceeds a pre-determined threshold, the camera node will detect moving objects within its monitoring area and sends the foreground/difference image to the base station for recognition and classification operations [13, 14].

2.1.1 OBJECT DETECTION USING THE BACKGROUND SUBTRACTION AND IMAGE RECONSTRUCTION

In most of the machine vision applications [23], detecting the moving objects in the camera nodes' field-of-view is one of the very important issues. To reduce the energy consumption of the network, Kenchannavar et al. [13] suggested that the camera nodes must process the captured image, and then send the useful information to the base station.

As shown in Figure 2, each camera node has a background image and periodically captures the images from its monitoring area. The moving objects are detected using a background subtraction technique with some pre-determined threshold values. If the background subtraction result exceeds the pre-determined threshold value, the camera node will detect moving objects within the monitoring area and calculates the energy that required for subtraction. Then, the difference image is sent to the base station using TCP/IP protocol. The foreground image is reconstructed using the background image, difference image and the median filter [24] at the server and the energy consumed during reconstruction and transmission is calculated.

![Block Diagram for the Camera node and server](image)

Figure 2: Block Diagram for the Camera node and server [13]

The results of authors' implementations show that the visual sensor network’s bandwidth [25] in the cases that the captured images have been processed is efficient as compared to without processing the images, which in turn increase the network's lifetime [26]. However, in subtraction/reconstruction approaches, non-object pixels with zero value (black) are injected into the network, when the background subtraction result is sent to the base station. Therefore, the transmission energy of the network is increased and its performance will not be acceptable. Furthermore, background subtraction-based object detection methods are sensitive to environmental factors like luminance [18].
2.1.2 MULTI-OBJECT DETECTION USING THE HAAR-LIKE FEATURES

To accelerate multi-object detection in VSNs with low training data, Vaidehi et al. [14] have been proposed a method using Haar-like features [27] and Joint-Boosting algorithm [28]. The foundation of work is that each camera node has a background image and periodically captures the foreground images from its field-of-view. If the difference between the background and foreground images exceeds a pre-determined threshold, foreground image is sent to the base station for extra recognition tasks.

As shown in Figure 3, the foreground images received are treated as the inputs to the multi-object detection system in the base station. In first step of the multi-object detection approach, the integral images [29] are defined to compute the Haar-like features fast. The integral image is a matrix in which its parts include sum of all the pixels in the left-upper part of the foreground image. The rectangle areas and their Haar-like features are directly obtained in the integral image. Haar-like features are the input of the decision tree classifiers. In the next step, after a strong classifier is trained, it can be applied to the foreground image's regions for detecting the object/non-object sub-window.

The simulation results of the author's implementation proved that classifier requires low training data by using Haar-like features, which in turn accelerate multi-object detection. Furthermore, the paper shows that detection system recognizes all instances of the objects regardless of their scale and location. However, further investigation states non-object pixels are injected into the visual sensor network, when all the foreground image's pixels are sent to the base station. Therefore, the VSNs' lifetime is decreased and proved inefficiency of the multi-object detection method.

2.2. OBJECT DETECTION METHODS WITH THE CAMERA NODES INVOLVEMENT

It is worth mentioning that calculating the difference between the background and foreground images are the basis task of all object detection methods in visual sensor networks. However, analyzing the conventional object detection methods shows using only the background subtraction techniques increases the network's energy consumption, which in turn decrease the lifetime of the camera nodes. To reduce data transmission cost in VSNs, some of the existence approaches involves the camera nodes to perform various preprocessing tasks after background subtraction and transmit only the useful information such as bounding box, features or key-points of the objects to the base station [16, 17, 18, 19].
2.2.1 Extracting Bounding Box of the Objects

To detect objects using minimum hardware and improve the detection costs, Pham et al. [16] have suggested an efficient approach for extracting the bounding box of the objects in camera nodes. In bounding box detection method, each stable camera node captures a background image and converts to gray-scale one. Then, the camera nodes periodically capture foreground images from own field-of-view and convert to gray-scale ones, too. The moving objects are detected using a background subtraction technique with some pre-determined threshold values. If the difference between background and foreground images exceeds the pre-determined threshold, the result of background subtraction that is a black-and-white image with only values of 0 and 1, will be treated as the input of the bounding box extraction step. As shown in Figure 4, bounding box extraction step involves row and column scans to detect whether the number of consecutive differences is greater than a pre-determined threshold or not (difference threshold for objects' length and width). During row scans, the first pixel location of the first threshold consecutive differences and the last pixel location of the last threshold consecutive differences are recorded as the row edges of bounding box. Similarly, in column scans, the first pixel location of the first threshold consecutive differences and the last pixel location of the last threshold consecutive differences have been recorded as the column edges of bounding box. The extracted rows and columns' edges are used for determining bounding box of the objects in the colored foreground image.

![Figure 4: Extracting the Bounding Box of Objects: (A) Black-and-white Subtraction Result, (B) Rows and Columns Scan, (C) Bounding Box of the Objects [16]](image)

It is worth mentioning that the noises in the image cause finding some false targets in the most of the object detection approaches [30]. The results of the authors' implementation show that images' noises are usually in small groups of consecutive pixels and the approach discards them. Furthermore, the paper proves that the transmission costs in visual sensor networks are more than processing ones. So, preprocessing images in the camera nodes and sending only the bounding box of the objects to the base station increase the network lifetime, significantly. However, further investigation shows extracting each object's bounding box eliminates non-object pixels in the overall bounding box and decreases the injected traffic into the VSNs. In other words, bounding box extraction approach can be suitable for only single object detection. Furthermore, in the human detection applications, extracting the bounding box of each face [31] and sending it to the base station is adequate to satisfy the recognition tasks requirements.

To further improvement in visual sensor networks' lifetime, some works have been proposed various low-complexity face detection methods in camera nodes for detecting the existence faces in the extracted bounding box. Yousefi et al. [32] have suggested an energy-aware multi-object method that works based on extracting the bounding box of the objects and Boosting-based face detection algorithm. The simulation results demonstrated that face detection method injects low volume of traffic into the network and saves camera nodes energy. However, the complexity of the Boosting-based face detection algorithm depends on the size of the input boxes and non-object pixels as the detected objects increases the size of the input boxes, which in turn raise the
processing complexity in the camera nodes. In the other hands, face detection methods are suitable only for the VSNs' applications with the aim of human recognition.

2.2.2 Extracting Objects' Features or Key-points

Some object detection approaches in visual sensor networks involve the camera nodes to perform preprocessing tasks, after background subtraction and sent only the features or key-points of the objects to the base station [17, 18, 19]. The qualities of the detected objects are accelerated by using the features or key-points extraction methods, especially in the applications that the distance between the camera node and objects is changing, continuously. Furthermore, extracting the features or key-points of the objects and injecting them into the networks decrease the processing and transmission costs, these in turn improve the performance of the visual sensor networks.

2.2.2.1 Object Detection Methods Based of Binary Robust Invariant Scalable Key-points

To maximize the quality of reconstructed pixel-domain representation under limited resources such as bandwidth and processing power, some approaches have been suggested that camera nodes extract the main features required for object recognition using Binary Robust Invariant Scalable Key-points (BRISK) and send them to the base station for extra recognition analysis. Object detection approach based on BRISK processes the foreground image to recognize a number of salient key-points that correspond to very different pixels of the underlying image. Finally, descriptors detected from the foreground image are matched with a set of descriptors extracted from a database of reference images. Therefore, a ranked list with the most relevant results is returned.

To use the BRISK for detecting objects, Redondi et al. [17] have extended Binary Robust Independent Elementary Features (BRIEF) designing to fix it against scale and rotation transformations. As shown in Figure 5, the camera nodes are responsible for capturing images, performing key-point detection tasks and finally, transmitting the descriptors to the base station. The base station performs object recognition leveraging the descriptors received from the camera node. The relay nodes only perform information communication and routing tasks.

![Figure 5: Architecture of Object Detection by BRISK [17]](image)

Paper's simulation results prove that the processing time depends on the images resolution and the number of highlight key-points. Furthermore, the paper shows object detection accuracy and data
transmission duration is accelerated by using the BRISK to detect the objects. However, using the local features such as BRISK for object detection in visual sensor networks has not a perfect performance to predict and formulate the lost data.

To solve the control problem and balance the processing loads in a visual sensor network, some works have used the temporal correlation in video sequences [19]. The basis of the work is distributing the processing load by allocating sub-areas of the images to the camera nodes. Therefore, the threshold and cut-point are estimated for each image, and then the optimal values of parameters via autoregressive models are predicted. The analytical results of the paper show that prediction-based methods reach the detection threshold and cut-points in a persuasive performance. Furthermore, achieving low computational complexity make them a group of convenient ways to control and balance the processing load on the local feature detection in VSNs. However, further investigations prove that prediction-based methods will not have an acceptable performance in visual sensor networks' monitoring and surveillance applications.

2.2.2.2 OBJECT DETECTION METHODS BASED ON ADAPTIVE GAUSSIAN MIXTURE MODEL

To eliminate the influence of environmental factors such as brightness variations, some works [18] have proposed object detection methods in visual sensor networks based on adaptive Gaussian mixture model. As shown in Figure 6, the first step of object detection in the camera nodes is frame reconstruction, which reduces the image size by partitioning it into blocks with a pre-determined size. Then, average color value of each image block is computed in a RGB color space and replaced for the corresponding block. In the second step, background modeling is performed. The static pixels may only be modeled by a Gaussian member, while other pixels, which are non-static, should be modeled by multiple Gaussian mixture components. Finally, to detect whether a pixel matches a component of the Gaussian, sort the Gaussian components, and then compare them one by one with the corresponding pixel. The pixels are detected as background if matches a component of the Gaussian and foreground one, otherwise.

Figure 6: A. Foreground Image, B. Frame Reconstruction, C. Object Detection in Reconstructed Image and D. Object Detection in Foreground Image.

Simulation results of the paper show adaptive Gaussian mixture model minimizes the processing cost and the influence of environmental factors. So, it is suitable for many detection-based applications in VSNs. However, further investigations prove the faces information is adequate for detecting the objects, which are humans. Therefore, it is proved transmitting the non-face information of the humans into the network reduces its lifetime.

3. CONCLUSION AND OPEN ISSUES

In this paper, the importance of object detection and recognition in visual sensor networks was discussed, which is set up in the monitoring and surveillance applications. Many different approaches have been proposed to solve the object detection problem in VSNs, considering
diverse metrics such as reliability, energy consumption, object detection accuracy and being real-time. In this paper, a survey on the object detection methods in visual sensor networks is presented for the first time. Two principal object detection categories in VSNs that explored in this paper were conventional object detection methods and object detection approaches with the camera nodes involvement. To be more precise, presented survey promoted an overview of recent object detection methods' literature with their performance evaluation.

Table 1 illustrates a summary of approaches discussed in this paper utilizing purpose and their cons and pros. It clearly indicates object detection approaches with the camera nodes involvement increase the preprocessing and transmission costs partially, which in turn accelerate lifetime of the camera nodes. Furthermore, table 1 shows that sending the face's information of the objects to the base station is adequate, when the objects are humans.

One of the important open research issues for object detection-based applications in visual sensor networks include decreasing the transmission energy of the networks and raising their lifetime. This purpose is achieved by increasing low cost preprocessing tasks in camera nodes and sending only the useful information (such as only the faces' information of the humans) to the base station. Another interesting issue for object detection approaches in visual sensor networks is energy efficient missing objects recovery, when the camera nodes fail. Therefore, fault tolerant object detection methods must be designed for VSNs to maximize the detection accuracy with minimum hardware.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Purpose of Methods</th>
<th>Advantage(s)</th>
<th>Disadvantage(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Detection using background subtraction techniques [13, 33]</td>
<td>Detecting the moving object</td>
<td>bandwidth usage reduction</td>
<td>Influencing of environmental factors, High energy consumption</td>
</tr>
<tr>
<td>Object detection using Haar-like features [14]</td>
<td>Detecting objects without size limitation</td>
<td>high-precision object detection, speed acceleration</td>
<td>High transmission energy</td>
</tr>
<tr>
<td>Extracting the bounding box of the objects [16]</td>
<td>High speed object detection with minimum hardware</td>
<td>Increasing residual energy of the network</td>
<td>Not perfect in reducing the transmission costs</td>
</tr>
<tr>
<td>Extracting the face's information of the objects [32, 34]</td>
<td>Increasing preprocessing tasks in camera nodes</td>
<td>Decreasing injected traffic into the network, Increasing network' lifetime</td>
<td>high processing and transmission costs</td>
</tr>
<tr>
<td>Object detection using BRISK [17, 35]</td>
<td>Maximizing the quality of pixel-domain display by limited resources</td>
<td>Optimizing processing time, Increasing detection accuracy</td>
<td>Inefficiencies in the prediction of lost information</td>
</tr>
<tr>
<td>Object detection by adaptive Gaussian mixture model [18]</td>
<td>Decreasing costs, Eliminating environmental factors</td>
<td>Reliable object detection</td>
<td>Sending objects' information to the base station instead of faces' one</td>
</tr>
</tbody>
</table>
REFERENCES


Authors

Shamim Yousefi received the B.S. degree in Information Technology Engineering and M.S. degree in Computer Engineering (Software) from University of Tabriz, Tabriz, Iran in 2013 and 2015, respectively. Currently, she is working as a researcher at the Wireless Ad hoc and Sensor networks research Laboratory (WASL) in University of Tabriz. Her current interests include lightweight methods for object detection and recognition in visual sensor networks.

Samad Najjar Ghabel is a lecturer at University of Mohaghegh Ardebili, Ardebil, Iran. He received the B.S. degree in Computer Engineering (Software) from University of Mohaghegh Ardebili, Ardebil, Iran and M.S. degree in Computer Engineering (Software) from University of Tabriz, Tabriz, Iran in 2013 and 2015, respectively. His main interests are Visual sensor networks, Computer Networks, networks security, developing and modeling software.