J. Vis. Commun. Image R. 25 (2014) 864–873

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Feature compression: A framework for multi-view multi-person tracking in visual sensor networks

Serhan Coşar *, Müjdat Çetin

Sabancı University, Faculty of Engineering and Natural Sciences, Orta Mahalle, Üniversite Caddesi No: 27, 34956 Tuzla-İstanbul, Turkey

ARTICLE INFO

Article history:
Received 14 May 2013
Accepted 30 January 2014
Available online 15 February 2014

Keywords:
Visual sensor networks
Camera networks
Human tracking
Decentralized tracking
Communication constraints
Feature compression
Compressing likelihood functions
Bandwidth-efficient tracking

ABSTRACT

Visual sensor networks (VSNs) consist of image sensors, embedded processors and wireless transceivers which are powered by batteries. Since the energy and bandwidth resources are limited, setting up a tracking system in VSNs is a challenging problem. In this paper, we present a framework for human tracking in VSNs. The traditional approach of sending compressed images to a central node has certain disadvantages such as decreasing the performance of further processing (i.e., tracking) because of low quality images. Instead, we propose a feature compression-based decentralized tracking framework that is better matched with the further inference goal of tracking. In our method, each camera performs feature extraction and obtains likelihood functions. By transforming to an appropriate domain and taking only the significant coefficients, these likelihood functions are compressed and this new representation is sent to the fusion node. As a result, this allows us to reduce the communication in the network without significantly affecting the tracking performance. An appropriate domain is selected by performing a comparison between well-known transforms. We have applied our method for indoor people tracking and demonstrated the superiority of our system over the traditional approach and a decentralized approach that uses Kalman filter.

1. Introduction

With the birth of wireless sensor networks, new applications are enabled by large-scale networks of small devices capable of (i) measuring information from the physical environment, such as temperature, pressure, etc., (ii) performing simple processing on the extracted data, and (iii) transmitting the processed data to remote locations by also considering the limited resources such as energy and bandwidth. More recently, the availability of inexpensive hardware such as CMOS cameras that are able to capture visual data from the environment has supported the development of Visual Sensor Networks (VSNs), i.e., networks of wirelessly interconnected devices that acquire video data.

Using a camera in a wireless network leads to unique and challenging problems that are more complex than the traditional wireless sensor networks might have. For instance, most sensors provide measurements of temporal signals that represent physical quantities such as temperature. On the other hand, at each time instant image sensors provide a 2D set of data points, which we see as an image. This richer information content increases the complexity of data processing and analysis. Performing complex tasks, such as tracking, recognition, etc., in a communication-constrained VSN environment is extremely challenging. With a data compression perspective, the common approach is to compress images and collect them in a central unit to perform the tasks of interest. In this strategy, the main goal is to focus on low-level communication. The communication load is decreased by compressing the raw data without regard to the final inference goal based on the information content of the data. Since such a strategy will affect the quality of the transmitted data, it may decrease the performance of further inference tasks. In this paper, we propose a different strategy for decreasing the communication that is better matched to problems with a defined final inference goal, which, in the context of this paper, is tracking.

There has been some work proposed for solving the problems mentioned above. To minimize the amount of data to be communicated, in some methods simple features are used for communication. For instance, 2D trajectories are used in [1]. In [2], 3D trajectories together with color histograms are used. Hue histograms along with 2D position are used in [3]. Moreover, there

* Corresponding author. Fax: +90 216 483 9550.
E-mail addresses: serhancosar@sabanciuniv.edu (S. Coşar), mcetin@sabanciuniv.edu (M. Çetin).

http://dx.doi.org/10.1016/j.jvcir.2014.02.004
1047-3203/

are decentralized approaches in which cameras are grouped into clusters and tracking is performed by local cluster fusion nodes. This kind of approaches have been applied to the multi-camera target tracking problem in various ways [4–6]. For a nonoverlapping camera setup, tracking is performed by maximizing the similarity between the observed features from each camera and minimizing the long-term variation in appearance using graph matching at the fusion node [4]. For an overlapping camera setup, a cluster-based Kalman filter in a network of wireless cameras is proposed in [5,6]. Local measurements of the target acquired by members of the cluster are sent to the fusion node. Then, the fusion node estimates the target position via an extended Kalman filter, relating the measurements acquired by the cameras to the actual position of the target by nonlinear transformations.

Previous works proposed for VSNs have some handicaps. The methods in [1–3] that use simpler features may be capable of decreasing the communication, but they are not capable of maintaining robustness. In order to adapt to bandwidth constraints, these methods choose to change the features from complex and robust to simpler but not so effective ones. As in the methods proposed in [4–6], performing local processing and collecting features to the fusion node may not satisfy the bandwidth requirements in a communication-constrained VSN environment. In particular, depending on the size of image features and the number of cameras in the network, even collecting features to the fusion node may become expensive for the network. In such cases, further approximations on features are necessary. An efficient approach that reduces the bandwidth requirements without significantly decreasing the quality of image features is needed.

In this paper, we propose a framework that fits with energy and bandwidth constraints in VSNs. It is capable of performing multi-person tracking without significant performance loss. Our method is a decentralized tracking approach in which each camera node in the network performs feature extraction by itself and obtains image features (likelihood functions). Instead of directly sending likelihood functions to the fusion node, a block-based compression is performed on likelihoods by transforming each block to an appropriate domain. Then, in this new representation we only take the significant coefficients and send them to the fusion node. Hence, multi-view tracking can be performed without overloading the network. The main contribution of this work is the idea of performing goal-directed compression in a VSN. In the tracking context, this is achieved by performing local processing at the nodes and compressing the resulting likelihood functions which are related to the tracking goal, rather than compressing raw images. To the best of our knowledge, compression of likelihood functions computed in the context of tracking in a VSN has not been proposed in previous work.

We have used our method within the context of a well-known multi-camera human tracking algorithm [7]. We have modified the method in [7] to obtain a decentralized tracking algorithm. In order to choose an appropriate domain for likelihood functions, we have performed a comparison between well-known transforms. A traditional approach in camera networks is transmitting compressed images. Both by qualitative and quantitative results, we have shown that our method can work under VSN constraints without degrading the tracking performance and it is better than the traditional approach of sending compressed images and a decentralized Kalman filter approach similar to the method in [5].

In Section 2, how we integrate multi-view information in our decentralized approach is described. Section 3 presents our feature compression framework in detail and contains a comparison of various domains for likelihood representation. Experimental setup and results are given in Section 4. Finally in Section 5, we conclude and suggest a number of directions for potential future work.

2. Multi-camera integration

2.1. Decentralized tracking

In a traditional setup of camera networks, which we call centralized tracking, each camera acquires an image and sends this raw data to a central unit. In the central unit, multi-view data are collected, relevant features are extracted and combined, finally, using these features, the positions of the humans are estimated. Hence, integration of multi-view information is done in raw-data level by pooling all images in a central unit. The presence of a single global fusion center leads to high data-transfer rates and the need for a computationally powerful machine, thereby, to a lack of scalability and energy efficiency. Compressing raw image data may decrease the communication in the network, but since the quality of images drops, it might also decrease the tracking performance. Thus, centralized trackers are not very appropriate for use in VSN environments. In decentralized tracking, there is no central unit that collects all raw data from the cameras. Cameras are grouped into clusters and nodes communicate with their local cluster fusion nodes only [8]. Communication overhead is reduced by limiting the cooperation within each cluster and among fusion nodes.

After acquiring the images, each camera extracts useful features from the images it has observed and sends these features to the local fusion node. The processing capability of camera nodes in emerging VSNs enable feature extraction at the camera nodes without the need to send the images to the central unit [9–12]. Using the multi-view image features, tracking is performed in the local fusion node. Hence, we can say that in decentralized tracking, multi-view information is integrated in feature-level by combining the features in small clusters. This both reduces the communication in the network and removes the need of powerful processors in the fusion node.

The decentralized approaches fits very well to VSNs in many aspects. The processing capability of each camera is utilized by performing feature extraction at camera-level. Since cameras are grouped into clusters, the communication overhead is reduced by limiting the cooperation within each cluster and among fusion nodes. In other words, by a decentralized approach, feature extraction and communication are distributed among cameras in clusters, therefore, efficient estimation can be performed.

Modeling the dynamics of humans in a probabilistic framework is a common perspective of many multi-camera human tracking methods [7,13–15]. In tracking methods based on a probabilistic framework, data and/or extracted features are represented by likelihood functions, $p(y|x)$ where $y \in R^d$ and $x \in R^m$ are the observation and state vectors, respectively. In other words, for each camera, a likelihood function is defined in terms of the observations obtained from its field of view. In centralized tracking, of course, the likelihood functions are computed after collecting the image data of each camera at the central unit. For a decentralized approach, since each camera node extracts local features from its field of view, these likelihood functions can be evaluated at the camera nodes and they can be sent to the fusion node. Then, in the fusion node the likelihoods can be combined and tracking can be performed in the probabilistic framework. A flow diagram of the decentralized approach is illustrated in Fig. 1. Following this line of thought, we have converted the tracking approach described in Section 2.2 to a decentralized tracker as explained in Section 2.3.

2.2. Multi-camera tracking algorithm

In this section we describe the tracking method of [7], as we apply our proposed approach within in the context of this method in