Journal of Network and Computer Applications 57 (2015) 119–128

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Multiple feature points representation in target localization of wireless visual sensor networks

Wei Li [a,b,*], Jorge Portilla [b], Félix Moreno [b], Guixuan Liang [b], Teresa Riesgo [b]

[a] School of Aeronautics and Astronautics, Sichuan University, No. 24, South Section 1, Yihuan Road, 610065 Chengdu, China
[b] Centro de Electrónica Industrial, Universidad Politécnica de Madrid, Jose Gutierrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

Article history:
Received 9 October 2014
Received in revised form
23 May 2015
Accepted 20 July 2015
Available online 1 August 2015

Keywords:
Point correspondence
Target localization
Wireless visual sensor networks

ABSTRACT

This paper discusses the target localization problem in wireless visual sensor networks. Additive noises and measurement errors will affect the accuracy of target localization when the visual nodes are equipped with low-resolution cameras. In the goal of improving the accuracy of target localization without prior knowledge of the target, each node extracts multiple feature points from images to represent the target at the sensor node level. A statistical method is presented to match the most correlated feature point pair for merging the position information of different sensor nodes at the base station. Besides, in the case that more than one target exists in the field of interest, a scheme for locating multiple targets is provided. Simulation results show that, our proposed method has desirable performance in improving the accuracy of locating single target or multiple targets. Results also show that the proposed method has a better trade-off between camera node usage and localization accuracy.

1. Introduction

In wireless visual sensor networks (WVSNs), sensor nodes that are equipped with cameras have functionalities of capturing visual information about targets and delivering the visual data to a base station for further analysis and decision making. Thus, WVSNs is capable of various security and surveillance applications, such as public security, facilities surveillance and monitoring. For most of these applications, the users are interested not only in existence of targets, but also in the positions of the targets (Liu et al., 2010), because the positions could facilitate target detection, recognition and tracking.

The task of localization provides with coordinates of both sensors and targets in sensor works (Soro and Heinzelman, 2009). Thus, localization task contains self-localization of sensor nodes and target localization. In this paper, we focus on the problem of target localization while the locations of sensor are already known. Target localization is to estimate the location of a target in the world coordinate based on the visual information of camera nodes (Kulkarni, 2007). The problem of target localization is well studied in wireless sensor networks. The measurement techniques in sensor localization include angle-of-arrival (AOA) measurements, distance related measurements and received signal strength (RSS) measurements (Mao et al., 2011). The existing techniques of target localization cannot be applied in WVSN. For example, multi-target can be cooperatively tracking by the Markov chain Monte Carlo data association method (Jiang and Hu, 2013). However, this method cannot address the problem of target localization in WVSN due to the significant differences in information capturing and processing between visual sensors and binary sensors. Actually, target localization in WVSN faces great challenges. Firstly, image processing is in general costly to implement in local nodes (Ercan et al., 2006), because the capabilities of computing are limited in local nodes. Secondly, the bandwidth resources are also restricted in WVSNs. Thus, there are constraints to transmit a huge amount of visual data generated by cameras to a central node or a base station (Charfi et al., 2009). Thirdly, since the sensing capability of a camera is characterized by directional sensing, the location information of a target in the depth dimension is lost in an image. Fourthly, due to the cost limitation, visual nodes in WVSN are equipped with low-resolution optical sensors (Akyildiz et al., 2007). Thus, the accuracy of filtering and extraction of target's position relevant information cannot be guaranteed in local sensor level.

Vision-based surveillance by multiple cameras receives considerable attentions, since visual surveillance by multiple cameras will enlarge the area and information from multiple views can be used to solve many problems (Liu et al., 2010). For example, the accuracy of the target localization can be gradually improved by

* Corresponding author at: School of Aeronautics and Astronautics, Sichuan University, No. 24, South Section 1, Yihuan Road, 610065, Chengdu, China.
E-mail address: li.wei@uestc.edu.cn (W. Li).

http://dx.doi.org/10.1016/j.jnca.2015.07.013
1084-8045/

selecting the most informative cameras based on correlation functions (Dai and Akyildiz, 2009) and the properties of the overlap region of the target in images (Li and Zhang, 2012) until the required accuracy level of target state is achieved. However, the multiple cameras bring new problems. Finding the correlated points pair in different images of cameras is a very difficult task. Furthermore, the energy and wireless channel capacity are very limited in WVSN. As discussed above, WVSN is a kind of resource-limited networks in nature. It is desirable to balance the trade-off between the accuracy of localization and the resources of WVSN.

The motivation of our study is to use the visual data acquired from the camera nodes to accurately estimate the position of target in the world coordinate. In this paper, we provide a method that uses multiple feature points to represent targets, and then provide a statistical approach to find the most correlated image point pair from different cameras, in order to reach the goal of improving the accuracy of target localization. Note that we focus on 2-D target localization on the ground plane. We assume that the cameras are placed horizontally around a room, which is the most relevant case for many real world applications. Besides, this paper makes the following assumptions about the wireless visual sensor network. Firstly, the location and orientation of each camera node is known within a universal coordinate system. Once a node enters into the networks, its geographical position remains constant. Next, all of the cameras are well calibrated. Finally, all of the nodes are time synchronized.

The initial results of this research have been published in Li et al. (2011), where we briefly introduced our target localization algorithm. In this paper, we expand on that work by providing further insight on the representation of the target by multiple feature points in target localization. In Li et al. (2011), we mainly focus on the single target localization. In this study, we provide a scheme for multiple target localization. The problems of the corresponding target matching and the occlusion between targets are also addressed in multiple target localization without prior knowledge. Besides, we also expend the experiments in a practical environment. Our main contributions include the following: 1) we design a method of using multiple feature points to represent a target in the goal of decreasing the influence of measurement noises and errors on the localization accuracy; 2) we proposed a statistical approach to match the most correlated feature points in different cameras; 3) we provide a scheme to address the issue of multiple target localization without prior information. The rest of this paper is organized as follows. Section 2 briefly highlights the related work. Section 3 presents the standard geometrical epipolar model to compute the possible position of single target. Section 4 proposes the technique to represent a target by multiple feature points and the method for finding the correlated point pair by the statistical method. Section 5 studies the case when there is more than one target existing in the field of interest. Section 6 conducts experiments to validate and evaluate the effectiveness of our proposed method and conclusions are given in Section 7.

2. Related work

Recently, a lot of researches focus on visual sensor networks, but very limited works related to the target localization in WVSNs has been reported.

Farrell et al. (2009) present a system that uses two cameras to localize the node of wireless sensor networks, and then employs non-imaging sensors to estimate the location of targets. Liu et al. (2010) described the common procedure of collaborative single target localization in wireless visual sensor networks, which extracts one feature point and adapts the epipolar geometrical model to compute the target position in the world coordination. Dios et al. uses Maximum Likelihood technique to fuse the cameras' observations about the location of target. They assume that the target in image can be represented as a single feature point and the level of noise in the measurements is low (Dios et al., 2011). Unfortunately, the target's extraction and localization are always corrupted by measurement noises and errors in practice. Massey et al. (2007) proposes methods to implement target localization using camera networks. They discussed two methods of triangulation for determining a target's position in the global coordinate space, grid-based coordination and convex polygon intersection scheme. Their methods heavily rely on an accurate algorithm to define the size, shape, presence, and position of a target within an image. Oztarak et al. (2009) provide a object localization approach which requires the distance between the extracted object and the camera. Unfortunately, the distance information cannot be obtained beforehand in most practical cases. Teng et al. (2014) proposed a method for locating electronic identifiers by integrating the electronic and visual signals. They designed a match engine to find the correspondence between an object's electronic identifier and its visual appearance.

Kulkarni (2007) uses two higher resolution cameras with overlapping coverage to localize an object and compute its Cartesian coordinates. Funiak et al. (2006) proposed a localization algorithm to retrieve the locations and poses of ad-hoc placed cameras in a sensor network by tracking a moving object. In their work, they assume that the object appearing in the image of cameras can be represented by a point. Therefore, they do not need to consider the problem of target correspondence. Lin et al. has provided a framework of a camera networks for tracking object in overlapping and non-overlapping fields (Lin and Huang, 2011), but they need prior knowledge about the features of an object.

All these methods mentioned above can be regarded as epipolar geometry-based solutions. Epipolar geometrical method of computing the coordinate of target is well studied, but it must solve the point correspondence problem because it is very hard to find the correlated image point pairs. Most of the aforementioned works assume that image features required for localization have already been acquired and that the correspondence between the image features are known. Some researchers used Scale-Invariant Feature Transform (SIFT) to find feature point correspondences (Sun and Liu, 2011). Since SIFT needs the frame level computation with iterated Gaussian blur operations on images and the frame difference operations on blurred images for feature extraction, its implementation faces challenges of heavy computation, large memory storage and long computational latency (Chiu et al., 2013). Medeiros et al. (2008) and Kurillo et al. (2008) use a LED bar to realize feature points correspondence between cameras. Different from the mentioned works, we present a statistical approach to find the most correlated feature point pair for participating in the task of target localization, without any additional tools and prior knowledge of the target.

3. General target localization model

In WVSN, since multimedia content, especially video streams, requires transmission bandwidth that is orders of magnitude higher than that supported by currently available sensors (Akyildiz et al., 2007), it is crucial to perform as much local processing (such as compression, error protection, filter, feature extraction, etc.) as possible to reduce the amount of information that needs to be communicated to other nodes (Sanchez-Matamoros et al., 2009). In this study, if cameras capture any target, they will extract the target position information from the raw image at local sensor. Afterwards, they send the position information to the base station where