Information Fusion 39 (2018) 120–138

Contents lists available at ScienceDirect

Information Fusion

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

Distributed data association in smart camera network via dual decomposition

Wan Jiuqing*, Chen Xu, Bu Shaocong, Liu Li

Department of Automation, Beihang University, 37 Xueyuan Road, Haidian District, Beijing, China

ARTICLE INFO

Article history:
Received 4 October 2016
Revised 1 March 2017
Accepted 22 April 2017
Available online 24 April 2017

Keywords:
Data association
Smart camera network
Dual decomposition
Distributed algorithm

ABSTRACT

One of the fundamental requirements for pedestrian surveillance using smart camera network is the correct association of each person's observations generated on different camera nodes to the person's track. Recently, distributed data association methods that involve only local information processing on each camera node and mutual information exchanging between neighboring cameras have attracted many research interests due to their superiority in large scale applications. In this paper, we propose a new method that performs global data association in a distributed manner by fusing the appearance and spatio-temporal measurements of objects captured by all camera nodes in the entire network. Specifically, we formulate the data association problem in smart camera networks as an Integer Programming problem by introducing a set of linking variables, and propose two distributed algorithms, namely L-DD and Q-DD, to solve the Integer Programming problem using the dual decomposition technique. In our algorithms, the original Integer Programming problem is decomposed into several subproblems, which can be solved locally on each smart camera. Different subproblems reach consensus on their solutions in a rigorous way by adjusting their parameters iteratively based on the projected subgradient optimization. The proposed method is simple and flexible, in that (i) we can incorporate any feature extraction and matching technique into our framework to calculate the similarity between observations, which corresponds to the costs of links in our model, and (ii) we can decompose the original problem in any way as long as the resulting subproblem can be solved independently and efficiently. We show the competitiveness of our method in both accuracy and speed by theoretical analysis and experimental comparison with state-of-the-art algorithms.

1. Introduction

Many research interests arose in recent years regarding wide-area surveillance using networked smart cameras with non-overlapping Field of Views (FOVs). The camera nodes constituting the network are smart, that is, they are not only able to collect video data, but also capable of local computation and mutual communication. They usually work cooperatively to discover or understand the behavior of some objects of interest, such as pedestrians or vehicles, moving in the monitored region. One of the fundamental prerequisites for achieving these goals is the correct reconstruction of camera-to-camera trajectory of each object, or equivalently, grouping observations originated from the same object, which may be generated by different cameras at different time instants, into a single track. This problem is often referred to as data association in camera network [1,2], trajectory recovery [3], or camera-to-camera tracking [4–8].

However, data association in non-overlapping camera network is a rather challenging task. Firstly, the visual appearance features may be indistinctive, e.g. persons are difficult to distinguish from one another if they wear similar clothes. In addition, as the cameras are mounted at different sites with various view angles and lighting conditions, the appearance of an object may undergo large variations across disjoint camera views, thus resulting in different objects appearing more alike than that of the same object. This makes the performance of the methods based solely on appearance cues [9–12] far from satisfactory.

An important way to improve the accuracy of data association is to combine the appearance and spatio-temporal cues. Spatio-temporal cues include the location of exits and entrances between two cameras, direction of movement and the average time spent in travelling from one camera to another. The pioneer work in this direction was presented in [13], which was followed and enhanced by lots of authors [1–8,14,15]. Unlike the cases of objects tracking in a single view [16] or overlapping views [17–19], where

* Corresponding author.
E-mail addresses: wanjiuqing@buaa.edu.cn (W. Jiuqing), xchen0530@163.com (C. Xu), busc829@foxmail.com (B. Shaocong), liulizi123@sina.com (L. Li).

http://dx.doi.org/10.1016/j.inffus.2017.04.007
1566-2535/

smooth and accurate motion models are usually available, there are lots of difficulties in tracking across non-overlapping views. For example, the travelling time between the two disjoint cameras may vary greatly from person to person, and there may be unpredictable pauses in blind regions. Thus the spatio-temporal constraints are usually inaccurate. One consistent way to overcome this inaccuracy is to utilize the spatio-temporal information in the entire network. Instead of modelling spatio-temporal relationship between two cameras independently, the spatio-temporal observations made in the whole network are preferred to be considered simultaneously. However, it is hard to model the complex dependency between spatio-temporal measurements generated by a large number of cameras.

For large scale networks, it is unrealistic to transmit the original video data collected by all camera nodes to a central server for processing due to the limitation in communication bandwidth. Smart cameras can alleviate this problem by analyzing the video data locally and transmit only the extracted features to the central server where some data association algorithm is running. Nevertheless, data association is a computationally combinatorial problem. The central server will become overwhelmed quickly when the number of observations is increasing. Therefore, to fully exploit the computation and communication resources of the smart cameras, distributed data association algorithms, which involve only local computation on each camera and information exchanging between neighboring cameras, are more desirable than centralized ones. However, finding the global optimal data association solution based solely on local computation and communication is far from a trivial matter.

In this paper, we propose a new distributed data association method for smart camera network to partially overcome the challenges mentioned above. Following [13,14], we use both appearance and spatio-temporal information for data association inference. To model the dependency between observations generated by camera nodes in the entire network, we transform the data association problem into a constrained Integer Programming (IP) problem. Then we present two distributed algorithms for solving the IP problem by using the Dual Decomposition (DD) technique [20,21]. The main contributions of this work are summarized as follows:

(i) We formulate the data association problem in smart camera network as a constrained IP problem. Specifically, we assign a binary linking variable to each possible link between observation pairs to indicate if they belong to the trajectory of the same person or not. The collection of all linking variables constitutes the optimization variable of the IP problem. The linking variables do not take value arbitrarily. They must satisfy the constraint that each observation should not be included in more than one trajectory. We define two kinds of energy functions to measure the likelihood of different linking configurations. The first is a linear energy function which is calculated by adding up the similarities between adjacent observations in each trajectory; the second is a quadratic energy function which is capable of evaluating higher-order similarities of three-observation tracklets. We calculate the similarity measure using simple appearance and spatio-temporal models, but more advanced feature extraction and matching techniques [9–14] can be applied directly to our framework.

(ii) We derive two distributed algorithms for solving the above two energy minimization problems respectively using the DD technique, which decomposes a large problem into several solvable subproblems and enforces consistency of solutions on common variables between subproblems by adjusting parameters of each subproblem in a rigorous way. Specifically, we propose two algorithms: L-DD (Linear DD) and Q-DD (Quadratic DD). L-DD is used for solving the linear energy minimization problem. L-DD decomposes the original IP problem into $2K$ (two for each camera, K is the number of cameras in the network) linear assignment subproblems, which are solved independently on each smart camera by using the Hungarian algorithm. Q-DD is used for solving the quadratic energy minimization problem. Q-DD decomposes the IP problem into N (one for each observation, N is the number of observations) subproblems, which are solved on each camera by direct searching. In L-DD and Q-DD, the parameters of each subproblem are updated at each iteration by using the projected subgradient method that maximizes the lower bound of energy function and enforces the consistency on solutions between overlapping subproblems. Both L-DD and Q-DD can be implemented in a distributed manner, that is, it involves only local operations on each camera and mutual information exchanges between neighboring cameras. It is also worthy to note that the DD framework is flexible in that we can design new algorithm by considering other kinds of decomposition as long as the resulting subproblems can be solved on each camera efficiently.

(iii) We analyze the proposed algorithms theoretically and experimentally. For L-DD, we prove that the computed energy lower bound is guaranteed to reach the optimal value of the energy function. In other words, L-DD is guaranteed to find the global optimal solution for the linear energy model. For Q-DD, we prove that the optimal lower bound of the quadratic energy function provided by our method is the same as that provided by a recently proposed centralized data association algorithm [22] which is based on Lagrangian relaxation of the original problem. Finally, we demonstrate the effectiveness of our algorithms by applying them to several simulated and real datasets and comparing them with the state-of-the-art methods.

The rest of this paper is organized as follows: In Section 2 we discuss the related works. In Section 3 we describe the data association problem formally and transform it into IP problem. We define two kinds of energy models and discuss the calculation of model parameters. In Section 4 we introduce the DD framework and propose two algorithms, L-DD and Q-DD, for minimizing the linear and quadratic energy functions, respectively. We also discuss the property of the lower bounds given by our algorithms theoretically. In Section 5 we report the experimental results of our methods and compare them with some state-of-the-art methods. Conclusion and further research perspective are given in Section 6.

2. Related works

The problem of data association in camera networks is closely related to but different from two other kinds of problems: person re-identification [9] and multi-camera tracking [17]. Person re-identification aims to visually matching individuals disappearing from one camera view in one or more other views of potential substantial distances and time differences. Person re-identification is usually based solely on appearance information. Most of the works in this field focus on feature extraction [10], metric learning [11] and matching techniques [12]. The goal of multi-camera tracking is to estimate the state, usually the position, or shape and velocity in more complex cases, of the targets monitored by several cameras with (usually) overlapping FOVs. Typically a state evolution model is available, based on which a lot of algorithms, such as Hidden Markov Model [23], particle filtering [24] or various kinds of consensus filters [18,25,26] have been proposed. The problem of data association in camera networks can be viewed as a special