ARTICLE IN PRESS

Signal Processing: Image Communication ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Multi-marker tracking for large-scale X-ray stereo video data

Xiaoyan Jiang [a,*], Marcel Simon [b], Yuan Yang [c], Joachim Denzler [b]

[a] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China
[b] Computer Vision Group, Friedrich Schiller University of Jena, Jena, Germany
[c] School of Instrument Science and Engineering, Southeast University, Nanjing, China

ARTICLE INFO

Keywords:
Multi-object tracking
Data association
Directed acyclic graph
X-ray
Big data

ABSTRACT

Analyzing large amounts of video data is one of the key challenges in the trend towards big data. In the field of medical research, for example, to analyze infected cardiac movements, stereo X-ray sequences of beating animal hearts implanted with radiopaque markers are recorded. As manual annotation of exact marker positions in large-scale recordings is time-consuming and infeasible, research on automatic tracking of multiple markers is a crucial task. We propose an efficient two-stage graph-based data association approach to tackle this problem. Difficulties of the sequences like 2D occlusions, low contrast, inhomogeneous movement, and inaccurate detections, are considered in the framework. Reconstructed 3D observations are modeled and connected using a weighted directed acyclic graph to obtain tracklets with high confidence via shortest path extraction. Afterwards, tracklets are linked into longer tracks by a tracklet graph in a similar manner while global features are adopted. The approach is validated on eight X-ray cardiac datasets of beating sheep hearts with various diseases. Outperforming standard tracking approaches, *e.g.* particle filter, the experimental results show a high accuracy comparable to human experts and efficiency in the meantime. The proposed approach is generic and can be directly applied to other video data as well.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, the trend towards big data also reached the area of medical research. One crucial task in this field is the analysis of recorded large-scale video data using object tracking techniques. For example, medical researchers are interested in the effect of certain diseases on the movement of the heart. For this purpose, cardiac sequences are recorded by a biplanar X-ray acquisition system where implanted radiopaque markers in the heart reveal the movement during the cardiac cycles and relevant muscle activities. The accurate tracking of multiple markers is an important step towards understanding and treating heart-related diseases since it is capable of tracking distinct cardiac structures [1,2]. Fig. 1 shows an example of the system and sample images.

However, tracking of the radiopaque markers is challenging in complex scenes due to the following difficulties:

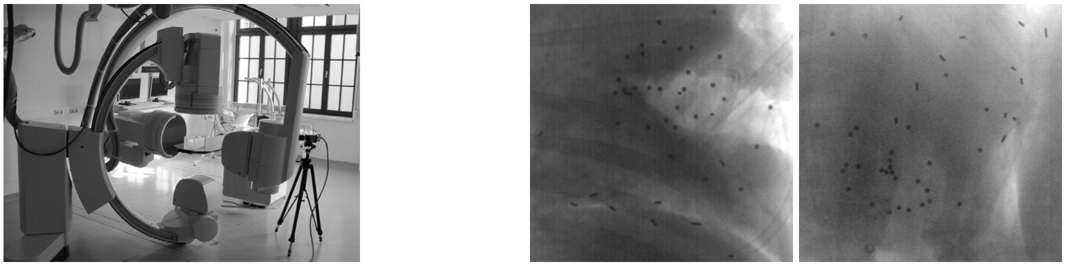
Occlusion Densely distributed objects (the markers) have inter-object occlusions and object-obstacle occlusions. As a result, appearance features of the objects are not temporally consistent, which is even worse if objects are occluded for a long period of time as it also commonly occurs in our application. This is a typical problem if only one single camera is used [3]. In our work, we hence use a biplanar camera setup and find matching markers in the two views by reconstructing the 3D positions.

Confusion The second most difficult case for visual tracking is "confusion" as pointed out by [4]. It happens when targets have indistinguishable appearance or similar-looking regions in the background are close to the object of interest. For instance, markers moving through anatomical structures. In this case, decision ambiguity may lead to mistakes. Others, *e.g.* bad observation viewpoints, small resolution images [5], background clutter, varying number of objects and inhomogeneous movement also confuse the tracking system.

Due to the improvement of detection algorithms [6,7] both in terms of accuracy and computational feasibility, tracking-by-detection (TbD) becomes a popular framework [8]. By continuously applying detectors to single frames, objects can be initialized automatically by the discrete detections. However, the output of detectors is sparse and unreliable and accuracy is far away from perfect since there are a lot of false positive and false negative detections (missing detections). Recursively searching for all possible associations of the detections is computational expensive and in many cases NP-hard. Therefore, efficient *data association* is crucial for the detection-based multi-object tracking and the goal is to robustly

* Corresponding author to: Xingzhen Lou, Room: 908, Longteng Road Nr. 333, Songjiang District, Shanghai 201620, China.
E-mail addresses: xiaoyan.jiang@outlook.com (X. Jiang), marcel.simon@uni-jena.de (M. Simon), yangyuancsi@163.com (Y. Yang), joachim.denzler@uni-jena.de (J. Denzler).

http://dx.doi.org/10.1016/j.image.2017.06.009
Received 6 January 2017; Received in revised form 8 June 2017; Accepted 23 June 2017
Available online xxxx
0923-5965/© 2017 Elsevier B.V. All rights reserved.



(a) Acquisition system.

(b) Example images.

Fig. 1. (a) Biplanar high-speed X-ray acquisition system (Neurostar®, Siemens AG), (b) X-ray images of left and right camera views showing a sheep heart with implanted markers.

link detections into tracks. According to literature, a multi-object TbD system generally consists of three units [3,9]: object detection, data association, and state estimation. In this paper, we focus on data association.

We present a two-stage graph-based data association approach to analyze the stereoscopic X-ray recordings. The method follows a TbD paradigm using 3D detections reconstructed from 2D detections as input and 3D object trajectories as output. The problem of finding meaningful linking among 3D detections is expressed by a probabilistic formulation, which is mapped to the shortest path searching algorithm in two consecutive graphs. In details, first, confident tracklets (object trajectory fragments) are extracted from a 3D detection based graph using spatial and temporal restrictions. Second, complete trajectories are then generated by linking tracklets modeled by a tracklet graph in a similar way using features extracted from tracklets. The framework assembles 3D detections to form probabilistically reasonable 3D trajectories in complex scenes. Note that the proposed framework is independent from specific scenarios and can be directly applied to any kind of video data.

The paper is based on our previous work presented in [10]. Our contributions are an extension of this work with respect to the following aspects: (1) a theoretical modeling of two-stage graph-based data association by probabilistic formulation; This formulation optimizes multiple object trajectories using 3D detections of the full sequence data which is robust to occlusions; (2) a thorough discussion of feature selections. Local and global features are adopted in two sequential graphs, which allows to adapt to different levels of semantic information in different stages. Associations are made with higher confidence as the weight function represents the distance between nodes more reasonably; (3) a comprehensive analysis of the approach by testing on various videos and comparison with state-of-the-art trackers. Additionally, the approach is easy to implement and has a much lower computational time compared to other approaches. In contrast to previous work, we also considered the key characteristics of big data, *i.e.* high volume, high velocity, and high variety [11] by keeping speed and a generic framework in mind.

2. Literature review

Numerous works have presented solutions to tackle the problem of data association [12–14]. Classic data association approaches include Multiple Hypotheses Tracking (MHT) [12] and Joint Probability Data Association (JPDA) [13]. Possible origins of target measurements in MHT are accounted by a set of data association hypotheses. For every measurement at a certain time step, the probability for three possible cases are calculated. The cases are (i) the measurement belongs to a previous track, (ii) it represents a new target, (iii) it is a false detection. After several time steps, as many measurements as time steps are obtained and the probabilities of joint hypotheses are computed recursively. Due to complexity, the analysis is limited to only a few such steps [3]. In comparison with the calculation of posterior probabilities for single measurements, JPDA considers independence between objects and computes joint conditional probabilities for data association. However, the computational complexity grows exponentially with the number of targets.

In contrast, Hungarian algorithm [15] can be adopted to find the best local assignment with a run time cubical in the number of targets [3]. [8] used Hungarian algorithm in the hierarchical data association approach. Afterwards, [3] presented a greedy approach which is assumed to be sufficient in practice [16] compared with Hungarian algorithm to match detections with particle filtering in pairwise frames. In the prediction stage, a constant velocity motion model is used to propagate the particles and each detection-tracker pair has a gating function to decide the association score. Additionally, the observation in a new time step is mainly updated by the associated detection, a boosted target-specific classifier. In [17], bipartite graphs were employed for tracking locally. Multi-object tracking is formulated as finding strong local minima of a continuous energy function presented in [18], which extends the Kalman filter [19]. Multiple judgments based on likelihood functions were embedded into the weight functions. Their approach outperforms the integer linear programming-based tracker presented in [20].

Compared with these local maximum or minimum searching scheme, recent approaches based on global optimization aim to model the state probability and investigate the best posterior probability over the entire sequence. This avoids locally optimal solutions and prevents association of non-stationary false positive detections. Sample approaches are network flow-based schemes [20–22] and graph-based approaches [9,23,24]. These approaches model the multi-object tracking issue as a combinatorial optimization problem which can be solved in polynomial time [21,22,25]. Graph-based approaches model the observations with nodes in a graph, where costs are assigned to edges to denote supporting scores for associations between different nodes. Solutions for the multi-object tracking problem are then obtained by searching paths with minimum or maximum cost. The tracking performance of graph-based algorithms depends on the objective function, and more specifically on the assigned costs to the edges.

Due to noisy detections, direct linking among detections often fails in challenging situations, such as occlusions. Sometimes multiple objects even merge and split during this occlusion, which causes great difficulties in maintaining the correct identities of the objects. This problem also occurs when objects are crossing each other. Since global optimization using detections over the whole sequence is time-consuming and challenging in deciding the suitable searching depth, splitting the complete object trajectories into meaningful fragments depending on the data itself and then linking fragments is more effective. By this way, connections with highly confident scores are formed in earlier stage. Depending the connected short fragments, longer tracks can be further linked. Consequently, a lot of research separated the tracking process into several stages [8,9,26,27] which aimed at producing more reliable tracklets (short track segments) in the first step. Afterwards in a second step, short tracklets are linked further into longer and reliable tracks. The Markov Chain Monte Carlo data association in [28] is an example for this. [29] proposed a three-layered data association scheme to achieve multi-ball tracking in monocular sequences. Ball candidates are first filtered by a sliding window fitted with a dynamic model at

Please cite this article in press as: X. Jiang, et al., Multi-marker tracking for large-scale X-ray stereo video data, Signal Processing: Image Communication (2017), http://dx.doi.org/10.1016/j.image.2017.06.009.

Download English Version:

https://daneshyari.com/en/article/6941742

Download Persian Version:

https://daneshyari.com/article/6941742

Daneshyari.com