ARTICLE IN PRESS

Pattern Recognition ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Pattern Recognition

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

Adaptive fusion of particle filtering and spatio-temporal motion energy for human tracking

Huiyu Zhou [a,*], Minrui Fei [b], Abdul Sadka [c], Yi Zhang [d], Xuelong Li [e]

[a] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT3 9DT, United Kingdom
[b] School of Mechatronics Engineering and Automation, Shanghai University, China
[c] School of Engineering and Design, Brunel University, United Kingdom
[d] Institute of Automation, Chongqing University of Post and Telecommunication, China
[e] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 13 April 2013
Received in revised form
7 February 2014
Accepted 12 May 2014

Keywords:
Computer vision
Object tracking
Occlusion
Colour
Motion energy

ABSTRACT

Object tracking is an active research area nowadays due to its importance in human computer interface, teleconferencing and video surveillance. However, reliable tracking of objects in the presence of occlusions, pose and illumination changes is still a challenging topic. In this paper, we introduce a novel tracking approach that fuses two cues namely colour and spatio-temporal motion energy within a particle filter based framework. We conduct a measure of coherent motion over two image frames, which reveals the spatio-temporal dynamics of the target. At the same time, the importance of both colour and motion energy cues is determined in the stage of reliability evaluation. This determination helps maintain the performance of the tracking system against abrupt appearance changes. Experimental results demonstrate that the proposed method outperforms the other state of the art techniques in the used test datasets.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Object tracking can be defined as a process of establishing temporal coherent correlations between image features over consecutive frames according to their shape, appearance and distance information [1–6]. Applications of object tracking have been commonly found in video surveillance [7], sports analysis [8], human motion analysis [9] and human–computer interface [10–12]. In particular, research studies on video surveillance applications mainly address challenging issues such as low quality images, illumination variations and cluttered backgrounds [13–16]. Interesting scenarios reported in these studies consist of traffic monitoring, gate access, crowd control and transportation security [17,18].

A working tracking system normally contains two key modules: prediction and correction. In the prediction module, a new state is obtained using the estimates made when people analyse the previous image frame, and a hypothesis is generated using a dynamic model. The correction module introduces an improved measurement, obtained at the current image frame, which refines the prediction. This new measurement is used in the next step as a new prediction. The system then assigns trust (or confidence) values to the prediction and the observation, respectively. The final estimation (or track) will be a compromise between the two locations which may or may not be close to the correct location. Nevertheless, it is hard to determine which resource is more trustable than the others in a challenging circumstance. This is mainly due to the appearance change of the subject to be tracked, or distraction because of the background clutters.

To deal with this problem, a tracking system needs reliable observations that are invariant to rotations, partial occlusions and pose changes. For example, local features or salient points, e.g. colours, textures, shapes or gradients, are commonly used in classical tracking systems for reducing ambiguity. To handle complicated situations such as crowds, people have applied corner/edge detection and optical flow calculation onto the corresponding stage in order to improve the discriminative capability. Furthermore, multiple visual cues or features have been fused so as to maintain the consistency of the observations. For example, colour cues have been integrated with shape information for human tracking in real applications.

However, empirical evidence shows that this integration scheme can fail to work in the case where subjects get dresses in similar colours, and background clutters may cause excessive uncertainty and ambiguity in shape discrimination. To handle this problem, Adam et al. [19] proposed to use patches to describe an

* Corresponding author. Tel: +44 28 90971753; fax: +44 28 90971702.
E-mail addresses: h.zhou@ecit.qub.ac.uk (H. Zhou), mrfei@staff.shu.edu.cn (M. Fei), Abdul.Sadka@brunel.ac.uk (A. Sadka), zhangyi@cqupt.edu.cn (Y. Zhang), xuelong_li@opt.ac.cn (X. Li).

http://dx.doi.org/10.1016/j.patcog.2014.05.006
0031-3203/© 2014 Elsevier Ltd. All rights reserved.

Please cite this article as: H. Zhou, et al., Adaptive fusion of particle filtering and spatio-temporal motion energy for human tracking, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.05.006

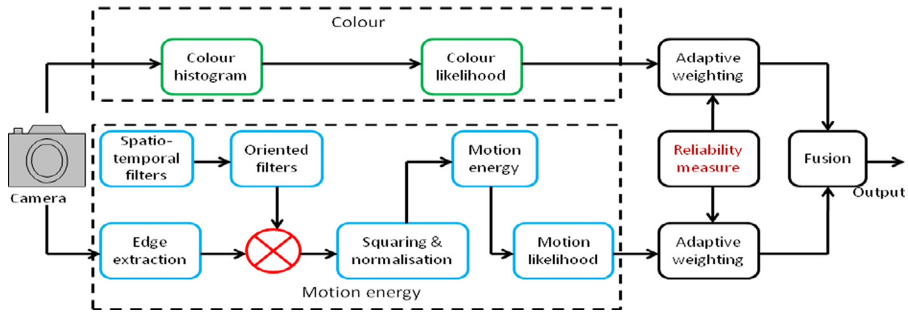

Fig. 1. Flowchart of proposed spatio-temporal based particle filter tracking system.

object, and then compared their histograms over the image frames. Babenko et al. [12] introduced an adaptive appearance model that embedded a discriminate classifier in an on-line manner to separate the object from the background. A low dimensional eigenbasis representation has been learnt, based on incremental principal component analysis algorithms [11]. In spite of promising outcomes in tracking accuracy, local feature based tracking systems have witnessed increasing challenges in practice (e.g. [20]). First of all, local feature based tracking systems require a selection of specific features extracted from the images. This subjective selection mechanism is operator dependent and may not effectively work in different environments. Secondly, it is a big challenge for one to estimate the centroid or velocity of a moving subject in the presence of occlusions.

In this paper, we introduce an adaptive feature fusion based tracking system to handle the problem of pose/illumination changes and occlusions. This scheme, despite being in its early stage of development, can be applied to event recognition and retrieval in the future incorporating the approaches such as [21]. Combining with the probabilistic data association algorithm e.g. [22], the proposed strategy can also be extended to the case of multiple person tracking. Fig. 1 shows the flowchart of the proposed system. We here propose to use an appropriate image representation, which refers to an effective fusion of local features (i.e. colour histograms) and the spatio-temporal motion energy [23]. The spatio-temporal representation can be used to capture the spatial appearance and dynamics of visual spacetime with certain immunity to appearance changes or clutters (this is different from the case reported in [24] that used colour and orientation cues). More importantly, when one of the fused features experiences fails, the other feature can be used to keep the systematic consistency and hence shows resilience to the complicated situations.

Compared to similar work such as [20,25,26], which take advantage of oriented energy features, multi-label and similar strategies, our major contributions are three-fold: (1) An adaptive fusion of the likelihoods associated with different features is presented in this paper. To be precise, motion energy cues are combined with colour histogram information within a particle filter based framework. It employs a reliability measure to determine the importance of colour and motion energy cues in the state estimation and updating, based on the historic colour or motion estimations. (2) The characteristics of the proposed scheme is investigated in terms of effectiveness and computational complexity. (3) A comprehensive evaluation of the proposed approach is performed against the other state of the art techniques.

This paper is organised as follows: Section 2 introduces the related work. Section 3 presents the proposed Spatio-tempOral Motion Energy Particle Filter (SomePF) tracking algorithm. This is followed by the description of the experimental work in Section 4. Finally, conclusions and future work are drawn in Section 5.

2. Related work

Prominent tracking systems developed to date include the Kalman filter [27] and the Condensation algorithm (also namely particle filter) [28]. Other existing techniques include mean shift [29], optical flow [30,31], multiple hypothesis tracking [32,33], Bayesian networks [34] and hidden Markov models [35].

Kalman filter has a long history to be intensively used in the community [27]. It is a state estimate method, based on linear dynamical systems that are perturbed by Gaussian noise. In most cases, a uni-model probability distribution is used in the state estimate. To release the linear assumption, a non-linear version, called extended Kalman filter (EKF), has been commonly used in object tracking [36]. However, the EKF also experiences some major issues in practice, one of which is that without the assumption of “static noise”, the estimated covariance matrix becomes unstable and hence causes the estimations drift away. Similar work like the iterated extended KF and unscented KF has also been reported in the community.

Particle filter is a conditional density propagation method that can be used to deal with non-Gaussian distributions and multi-modality cases [28]. This method allows a posterior distribution, estimated in the previous image frame, to be sampled with a set of particles. These particles are then propagated iteratively to successive frames using continuously updated observations and a prediction model. Particle filters have been popularly used to handle various tracking problems, e.g. [37–39]. However, the performance of particle filters can degrade as the dimensionality of the state space increases, and the support of the likelihood decreases (e.g. less sample numbers). One of the effective solutions is to combine a variational approximation with efficient importance sampling to achieve tracking recursion [39].

Mean shift is a technique of locating the maxima of a density function using the samples from that function [40]. This involves an iterative procedure, which starts with an initial guess and stops when the difference between two consecutive estimates is smaller than a pre-defined threshold. A kernel function is used to determine the weight of neighbouring points for re-estimation of the mean value. Mean shift has been used with colour histograms to find the peak of a confidence map for locating an object's position [29]. Yilmaz [41] introduced an asymmetric kernel mean shift, in which the scale and orientation of the kernel adaptively change depending on the observations at each iteration. Freedman and Kisilev [42] presented a novel fast mean shift procedure, based on random sampling of the Kernel density estimates. Yeh and Hsu [43] reported a feature selection algorithm used within a mean shift scheme, which adopted the AdaBoost algorithm to select features that best compensate each other and determine their weights using likelihood estimations. A comprehensive evaluation has been conducted in [44], which shows that mean shift based trackers have better performance than the variant CAMShift tracker.

Please cite this article as: H. Zhou, et al., Adaptive fusion of particle filtering and spatio-temporal motion energy for human tracking, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.05.006

Download English Version:

https://daneshyari.com/en/article/10360330

Download Persian Version:

https://daneshyari.com/article/10360330

Daneshyari.com