Pattern Recognition 61 (2017) 1–14

Contents lists available at ScienceDirect

Pattern Recognition

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

A collection of challenging motion segmentation benchmark datasets

Muhammad Habib Mahmood *, Yago Díez, Joaquim Salvi, Xavier Lladó

Computer Vision and Robotics (VICOROB) Institute, University of Girona, Campus Montilivi, Edifici P-IV, 17071 Girona, Spain

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form
28 June 2016
Accepted 3 July 2016
Available online 5 July 2016

Keywords:
Motion segmentation
Tracking
Trajectory
Benchmark
Dataset

ABSTRACT

An in-depth analysis of computer vision methodologies is greatly dependent on the benchmarks they are tested upon. Any dataset is as good as the diversity of the true nature of the problem enclosed in it. Motion segmentation is a preprocessing step in computer vision whose publicly available datasets have certain limitations. Some databases are not up-to-date with modern requirements of frame length and number of motions, and others do not have ample ground truth in them. In this paper, we present a collection of diverse multifaceted motion segmentation benchmarks containing trajectory- and region-based ground truth. These datasets enclose real-life long and short sequences, with increased number of motions and frames per sequence, and also real distortions with missing data. The ground truth is provided on all the frames of all the sequences. A comprehensive benchmark evaluation of the state-of-the-art motion segmentation algorithms is provided to establish the difficulty of the problem and to also contribute a starting point. All the resources of the datasets have been made publicly available at http://dixie.udg.edu/udgms/.

1. Introduction

The advances being done in computer vision are greatly dependent on the characteristics of the data bank of images and videos which algorithms use as benchmark. If the data encapsulates the true nature of the desired problem, then the proposed solution can be rigorously tested, results can be repeatable, standardization of evaluation measures can be achieved, and new algorithms can be adequately compared. However, if the bank of data does not envelop the correct nature of the task in question, or if a trivial subset of the problem is captured which is not representative of the actual problem, then the solutions proposed might be limited in nature, unstable in results, and thus might hamper further research. Due to these reasons, nowadays, the importance of datasets is immense as they are shaping the way forward for computer vision algorithms.

Motion Segmentation (MS) is one such research area, in which temporally continuous set of frames, called a sequence, is processed to provide a unique label to every motion present in those frames. It is a preprocessing step for several computer vision problems, i.e. semantic segmentation, crowd estimation, and surveillance [1–4]. MS can be performed on precomputed set of sparse trajectories of a sequence, or it can be directly applied on a desired set of frames to get motion region labels. In the huge amount of video data that is captured by hand-held devices these days, moving objects exhibit a variety of characteristics. MS datasets must be representative of these characteristic attributes of motion to capture the diversity of the problem.

The main limitations of the state-of-the-art of MS datasets, as seen in Fig. 1, are that they are made up of sets of videos where the use of synthetic sequences, like passively moving checker boards or other static objects, still prevails [5]. The motions in these datasets are captured in short sequences, with little or no background change. The number of motions per sequence is few, and even a video shot capturing 4 or 5 moving objects does not exhibit considerable amount of movement as is present in daily-life natural scenes. The results of algorithms are put forth on these databases because of the absence of a challenging, diverse dataset. In some cases, improved results are presented on less representative databases which do not help the cause. There are other small datasets, like Extended-Yale benchmark [7], SegTrack [8] and Cambridge CamShift dataset [9], which motion segmentation community has used to present their results. They are limited and not widely used because of fewer sequences or a smaller number of motions. A recently proposed dataset [6] tried to overcome these limitations, but the ground truth is provided only on 5% frames. Therefore, the need for a more diverse and challenging motion segmentation database, with which the borders of research in MS algorithms can be expanded, still persists.

Hence, in this paper we present our multifaceted diverse publicly available motion segmentation benchmark dataset of 39

* Corresponding author.
E-mail address: mhabib82@hotmail.com (M.H. Mahmood).

http://dx.doi.org/10.1016/j.patcog.2016.07.008
0031-3203/

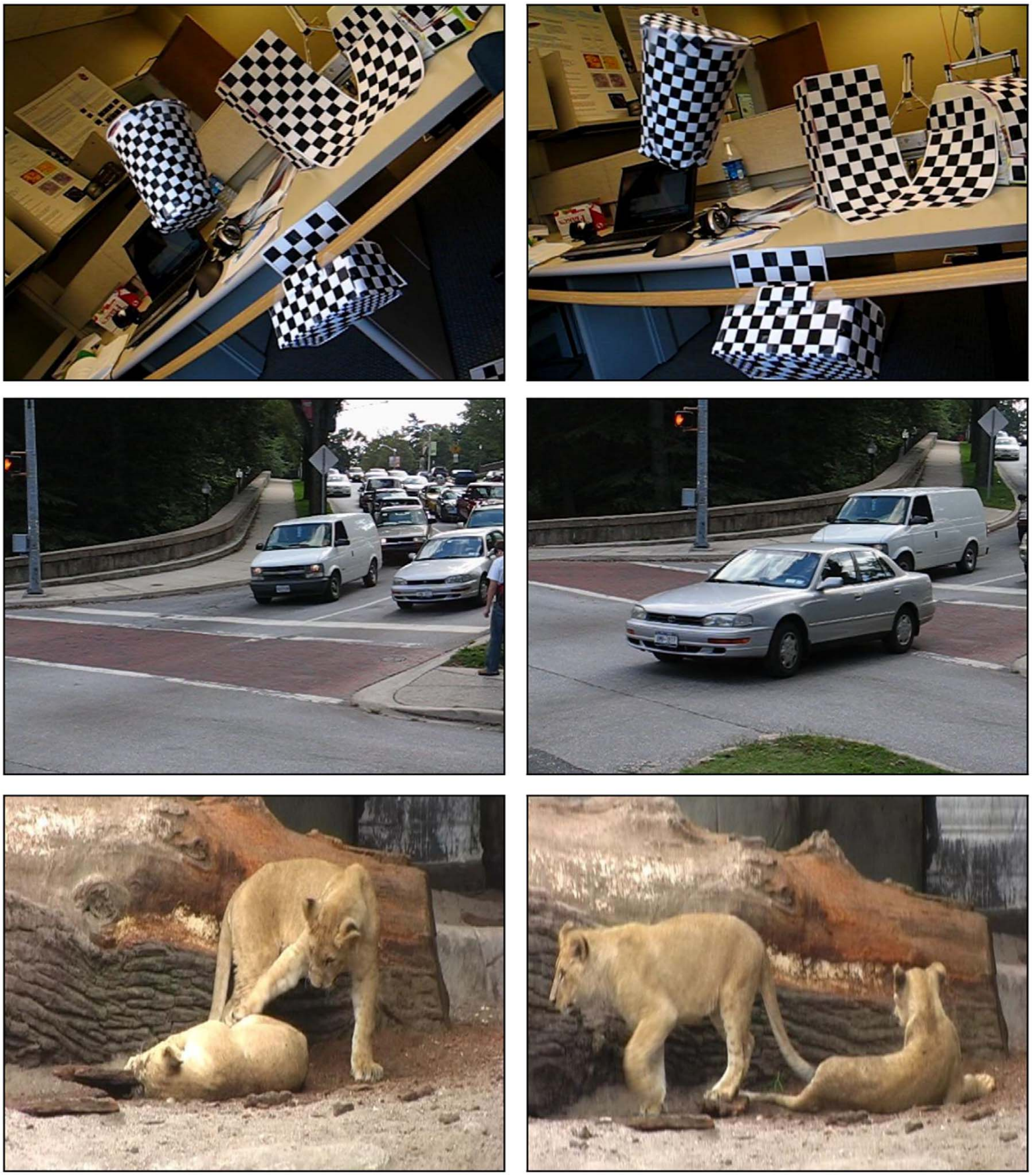

Fig. 1. First and last frame of video shots from the state-of-the-art in MS datasets. Top: '1R2TCRT', total 42 frames in Hopkins [5]. Center: 'cars9', total 60 frames in Hopkins and FBMS [6] both. Bottom: 'lion02' total 416 frames in FBMS.

long, and 312 short sequences with ground truth available in all the frames of all the sequences. The ground truth on 19 long, and 162 short sequences is provided as a trajectory label on all the trajectories of all the motions. And the ground truth on 20 long, and 150 short sequences is provided as a region label on all the motion regions of all the sequences. The average number of motions per sequence is almost 11, and the average frame length is around 815 frames. With a goal to overcome all the limitations present in the state of the art, all the sequences contain a fusion of real noise and distortions. The captured characteristics of noise in the sequences are missing data, partial/complete/multiple occlusion, stopping motion, multiple appearance–disappearance of objects, perspective distortion, etc. A further subset of 40 trajectory-based and 34 region-based short sequences with complete data is also provided to test algorithms that are unable to deal with missing data. Ochs–Brox (OB) [6] algorithm is used to provide a benchmark on long sequences as besides OB no other algorithm can be applied on these long sequences because of their complexity. Moreover, on short sequences a benchmark analysis with six well-known state-of-the-art MS algorithms, i.e. SSC [10], ALC [11], LRR [12], LS3C [13], ELSA [14], OB [6], is also presented for detailed quantitative and qualitative evaluation. The evaluation metrics encapsulate all the criteria used in the state-of-the-art datasets for an in-depth analysis. Moreover, a course of action on how to improve results on this database to produce robust algorithms is also suggested. A seminal work of this challenging dataset was presented in [15].