INFIERI2019@ HUST

Learning Dynamic Background for Moving Object Detection

Zhijun Zhang[1], Yi Chang[1] , Sheng Zhong[1] , Luxin Yan[1]
1. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

Introduction

Moving object detection in dynamic scenes is a challenging problem due to the complex variations in the background such as the swaying leaves and rippers. Most of the existing methods usually cast the variation into background or foreground layer, and thus fail to precisely model the dynamic variation in an explicit manner. In this work, we formulate the moving object detection as tensor-based three layers decomposition problem: static background, moving object foreground, and dynamic background to address this issue. Specifically, we additionally introduce a data-driven discriminative prior embedded into the maximum a posterior (MAP) framework to explicitly capture the dynamic background, which can be satisfactorily represented by a two stream spatiotemporal convolutional neural network. As for the static background, we introduce the low-rank tensor prior to model the temporal correlation; For the moving object, we utilize its spatiotemporal continuity via the Markov random field. Compared with previous matrix-based methods, each term in our method is tensor-based which could better preserve the intrinsic spatiotemporal structure correlation in the video. In addition, our model could be further improved by involving a semantic prior as an adaptive attention. The proposed method is extensively evaluated on several benchmarks, and significantly outperforms state-of-the-art methods.

Keywords: Moving Object Detection, Dynamic Background Variation, Clutter Modeling, MAP framework

Design and Implementation

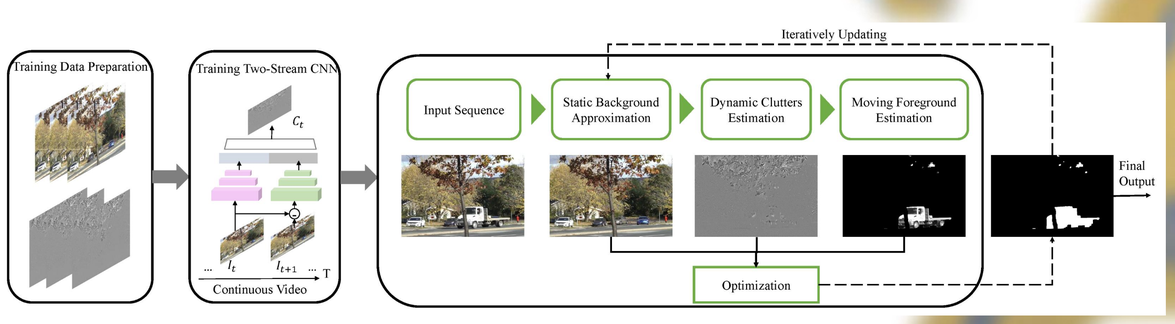

Fig. 1. Block design of the clock distribution The overview of the proposed approach. Our approach is mainly b ased on the MAP framework with integration of a data-driven discriminative prior for explicit dynamic clutter m odeling. The dynamic clutter is satisfactorily represented by a trained content-aware two-stream spatiotemporal CNN. Incorporated with spatiotemporal structure modeling of moving foreground (sparsity and smooth) and stat ic background (temporal low-rank), the unified framework can be solved by the alternating minimization algorit hm iteratively.

The Tensor Decomposition Model

The key problem of moving object detection from dynamic background is that the modeling of dynamic clutter is unreasonable. To address the problem, we propose a novel decomposition model that separates the dynamic clutter from the video. For a video sequence tensor $D \in \mathbb{R}^{H\times W\times T}$, the model is formulated as follow,

$$D = B + F + C + N$$

where $B, F, C \in \mathbb{R}^{H\times W\times T}$ denote static background, moving foreground, dynamic clutter and Gaussian noise tensor respectively.

Dynamic Background: Two-Stream CNN

We introduce a context-aware learning prior to model the distribution of dynamic clutter. CNN, as well known for its powerful representation capability, is a natural choice for the problem. Therefore, we design a two-stream spatiotemporal CNN to model the dynamic clutter.

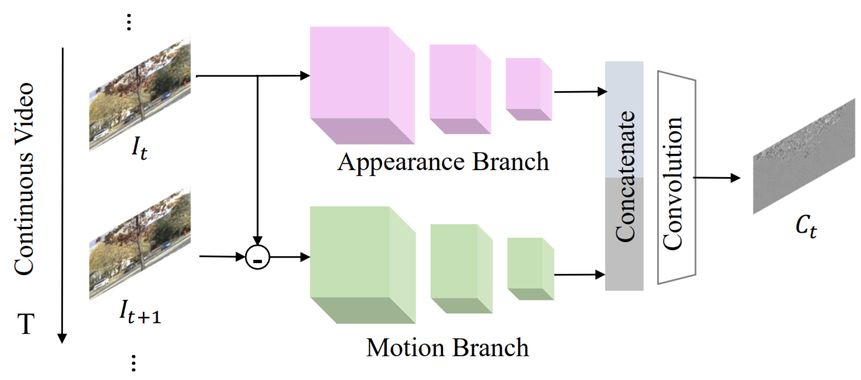

Fig.2. The overview of two-stream spatiotemporal CNN

Static Background: Low-rank Tensor

We model the static background in a low rank tensor. Notably, the low-rank property of static background tensor is anisotropy, where the spatial dimensions have no similar low-rank property with temporal.

Moving Foreground: Spatiotemporal MRF

For moving foreground, we assume a prior that foreground objects are contiguous and sparse in spatial and temporal space. Thus we use MRF to model the probability and correlations of foreground pixels.

Formulation & Optimization

Based on the modeling of dynamic clutter, static background and moving foreground aforementioned, our moving object detection method builds on Learning Dynamic Background (LDB), which can be formulated as:

$$\min_{B,C,S_{ijk}\in\{0,1\}} \frac{1}{2}\left\|P_{S^{\perp}}(D-B-C)\right\|_F^2 + \alpha rank_3(B) + \beta E(S) + \gamma\|C - f_{CNN}(D)\|_F^2$$

The first term in the equation is the data constrain respect to the decomposition model, i.e. video D_{ijk} is best fitted by static background B_{ijk} plus dynamic clutter C_{ijk} when no moving foreground S_{ijk} = 0. The second term forces that static background B is temporal unidirectional low-rank. The third term E(S) is the MRF energy function of S, which models the sparsity and spatiotemporal smoothness of moving foreground. The last term means the dynamic clutter can be well represented by constructed two-stream spatiotemporal CNN.

For optimization, we adopt an alternating algorithm that separates the energy minimization over B, C, and S into three steps.

Results and Conclusions

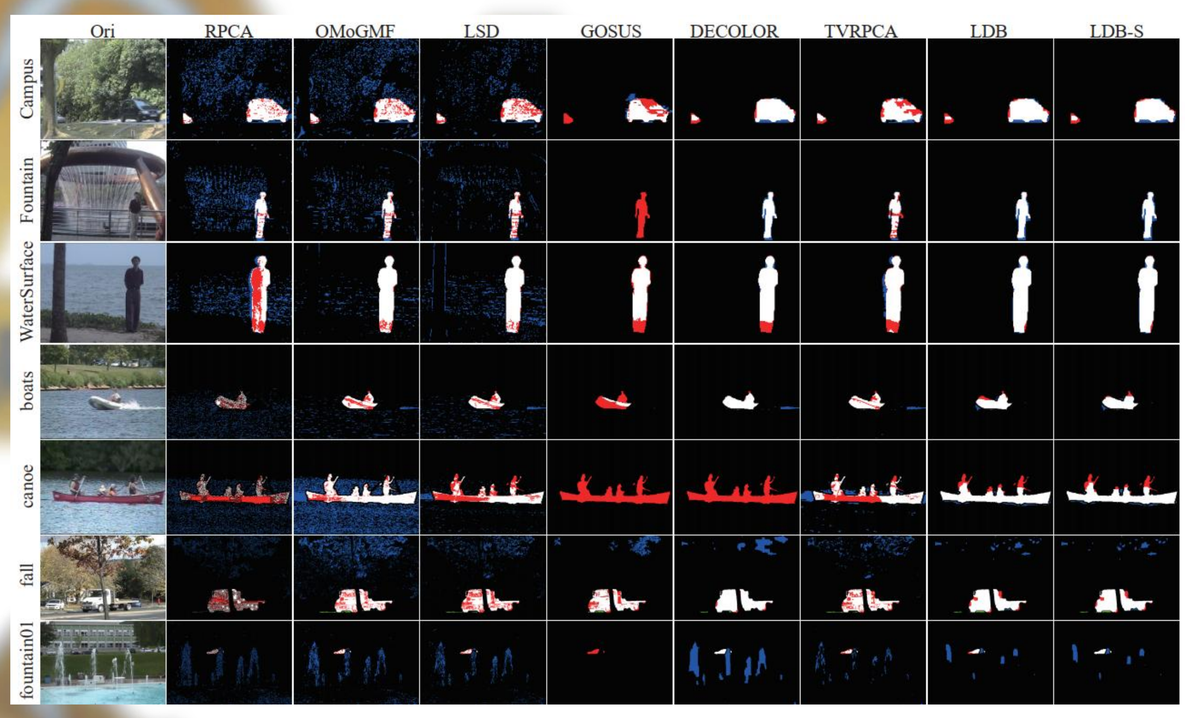

Fig.3. Comparison results on I2R and CDnet dataset

Video	dynamicBackground							badWeather				
	boats	canoe	fall	foun.01	foun.02	over.	Avg	blizz.	skat.	snowF.	wetS.	Avg
RPCA [4]	0.13	0.14	0.36	0.02	0.02	0.52	0.20	0.10	0.59	0.58	0.69	0.49
OMoGMF [35]	0.27	0.26	0.23	0.04	0.97	0.78	0.35	0.63	0.86	0.81	0.72	0.76
LSD [22]	0.35	0.47	0.35	0.04	0.09	0.72	0.34	0.70	0.85	0.80	0.74	0.77
GOSUS [33]	0.34	0.74	0.60	0.00	0.08	0.79	0.43	0.00	0.80	0.20	0.44	0.36
DECOLOR [41]	0.92	0.00	0.69	0.03	0.56	0.94	0.53	0.90	0.93	0.90	0.89	0.91
TVRPCA [6]	0.79	0.52	0.46	0.08	0.65	0.83	0.56	0.84	0.84	0.83	0.78	0.83
LDB	0.90	0.90	0.79	0.12	0.93	0.93	0.76	0.90	0.89	0.93	0.89	0.90
LDB-S	0.92	0.92	0.79	0.14	0.93	0.95	0.78	0.90	0.90	0.93	0.89	0.91

Table.1. Performance evaluation on CDnet dataset using F-measure. The average F-measures of the dynamicBackground and badWeather categories are shown in the eighth (Avg) and the last column (Avg) respectively. Red: best, blue: the second best.

In this paper, we propose a novel and unified model to address moving object detection from dynamic background. In this model, a video is decomposed into a unidirectional low-rank static background, a sparse and smooth moving foreground, as well as a CNN represented dynamic clutter. Embedded with the data-driven discriminative prior of dynamic clutter, the unified model can properly decouple the variations of background with moving foreground. In addition, involved with semantic prior of objects, the performance of method can be further improved. Experiment results show that the proposed methods outperform the state-of-the-art methods both qualitatively and quantitatively in the challenge datasets

Acknowledgment

The authors acknowledge the school of Artificial Intelligence and Automation of Huazhong University of Science and Technology for providing funding.