

12-13-2019

Moving Object Detection Based on Deep Convolutional Neural Network

Yuqiu Lu

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;

Jinyu Sun

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;

Shiwei Ma

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Moving Object Detection Based on Deep Convolutional Neural Network

Abstract

Abstract: *A deep convolution neural network MONet is designed for the intermittent motion problem in moving objects detection. In the absence of training data sets, a synthetic dataset GoChairs is generated by affine transformation, and on this basis the network training and testing are performed. The results show that the trained MONet can effectively detect the moving objects based on the correspondence between the pixels. The traditional datasets CDnet and I2R are also tested to verify the generalization performance of the network. In addition, MONet is compared qualitatively and quantitatively with classical methods for the intermittent motion problem of the objects. The experimental results demonstrate the superiority of the network in detecting the objects with intermittent motion.*

Keywords

moving object detection, intermittent motion, deep convolutional neural network, synthetic dataset

Recommended Citation

Lu Yuqiu, Sun Jinyu, Ma Shiwei. Moving Object Detection Based on Deep Convolutional Neural Network[J]. Journal of System Simulation, 2019, 31(11): 2275-2280.

基于深度卷积神经网络的运动目标检测方法

卢裕秋, 孙金玉, 马世伟*

(上海大学 机电工程与自动化学院, 上海 200444)

摘要: 针对运动目标检测中的间歇性运动问题, 设计了一个深度卷积神经网络 *MONet*。在缺乏训练数据集的情况下, 利用仿射变换生成一个合成数据集 *GoChairs*, 并在此基础上进行网络的训练和测试。结果表明, 训练后的 *MONet* 能够有效地根据像素点之间的对应关系检测出运动的目标。传统的运动目标检测数据集 *CDnet* 和 *I2R* 被用于测试以验证该网络的泛化性能。针对目标的间歇性运动问题, *MONet* 与经典方法进行了定性和定量的比较。实验结果证明了该网络在检测间歇性运动的目标时的优越性。

关键词: 运动目标检测; 间歇性运动; 深度卷积神经网络; 合成数据集

中图分类号: TP391.4 文献标识码: A 文章编号: 1004-731X (2019) 11-2275-06

DOI: 10.16182/j.issn1004731x.joss.19-FZ0368

Moving Object Detection Based on Deep Convolutional Neural Network

Lu Yuqiu, Sun Jinyu, Ma Shiwei*

(School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China)

Abstract: A deep convolution neural network *MONet* is designed for the intermittent motion problem in moving objects detection. In the absence of training data sets, a synthetic dataset *GoChairs* is generated by affine transformation, and on this basis the network training and testing are performed. The results show that the trained *MONet* can effectively detect the moving objects based on the correspondence between the pixels. The traditional datasets *CDnet* and *I2R* are also tested to verify the generalization performance of the network. In addition, *MONet* is compared qualitatively and quantitatively with classical methods for the intermittent motion problem of the objects. The experimental results demonstrate the superiority of the network in detecting the objects with intermittent motion.

Keywords: moving object detection; intermittent motion; deep convolutional neural network; synthetic dataset

引言

运动目标检测是计算机视觉领域的一个重要研究课题。其目的是利用视频帧序列之间的相关关系, 快速、准确地从视频中检测出运动的目标, 以

降低后续行为分析等环节的处理难度。在安防监控领域^[1], 基于运动目标检测技术的新型智能摄像头能够实现运动目标的实时检测与追踪。在智能交通领域^[2], 交通监管部门通过利用运动目标检测技术实时测量车辆的速度并对超速的车辆及时进行记录。在动作识别领域^[3], 运动目标检测作为特征提取前的预处理可以有效地减少计算量。在人机交互方面^[4], 基于运动目标检测技术的手势交互方式与依赖于键盘和鼠标的交互方式相比, 操作更加简



收稿日期: 2019-06-20 修回日期: 2019-07-23;
基金项目: 新疆兵团重大项目子项目(2018AA008-04);
作者简介: 卢裕秋(1994-), 男, 江苏宿迁, 硕士生, 研究方向为模式识别; 孙金玉(1991-), 女, 江苏连云港, 博士生, 研究方向为图像处理; 马世伟(通讯作者 1965-), 男, 甘肃嘉峪关, 博士, 教授, 研究方向为信号处理、模式识别。

<http://www.china-simulation.com>

• 2275 •

单。因此,对运动目标检测技术进行研究,具有重要的学术价值与应用价值。

目前,针对运动目标检测,许多学者提出了各种算法。然而,由于实际场景复杂多变,算法在应用时面临诸多难点,具体包括:1) 动态背景:背景中大多包含局部的扰动,比如树枝的摇曳、水面的波动。2) 目标阴影:运动目标的阴影会随着运动目标一起运动,因而很容易被算法误判为前景目标。3) 间歇性运动:实际场景中,物体通常存在运动与静止两种状态的转换。此外,随着社会智能化程度的加深,越来越多的应用对算法的实时性提出了更高的要求。

本文首先介绍了运动目标检测领域的经典算法以及研究人员在利用深度学习进行目标检测方面进行的探索。在此基础上,本文借鉴了经典光流网络 FlowNet 的思想,构建了一个端到端的深度卷积神经网络 MONet。该网络在一个人工合成数据集上进行训练和测试。此外,为了与经典的检测方法进行对比,网络也在 CDnet 和 I2R 两个数据集上进行了测试。结果表明,该网络具有较好的泛化性和较高的实时性,并且当场景中存在间歇性运动的目标时,该网络的性能要优于传统的方法。

1 相关方法

1.1 传统方法

作为计算机视觉领域的一个研究热点,研究人员提出了许多经典的算法。

Anderson 等^[5]提出 2 帧差分法。该算法复杂度低、运算速度快,但容易产生前景空洞。Wren 等^[6]根据光照变化的特点提出了单高斯模型(Single Gaussian model, SGM)。该算法将图像中各点的颜色值分布拟合单高斯函数。在此基础上,Stauffer 等^[7]提出高斯混合模型(Gaussian Mixture Model, GMM)以适应真实场景的多模态性。不论是 SGM 还是 GMM,都基于背景模型假设,但该假设并不一定符合实际。针对该问题,Elgammal 等^[8]提出核密度估计法(Kernel Density Estimation, KDE),

Barnich 等^[9]提出了 Vibe 算法。与上述方法不同,光流法通过计算像素点的瞬时速度来区分前景和背景。传统的光流计算方法包括: Horn-Schunck 光流法^[10]、Lucas-Kanade 光流法^[11]和金字塔 LK 光流法^[12]。另一类典型的运动目标检测方法是低秩稀疏模型法。Candes^[13]首先提出利用鲁棒主成分分析(Robust Principal Component Analysis, RPCA)恢复低秩矩阵,进而完成前景与背景的分割。之后,大量学者对 RPCA 算法进行了改进。Lucas^[14]提出了贝叶斯鲁棒主成分分析(Robust Principal Component Analysis, BRPCA)以提高模型参数 λ 的适应性; Bouguet^[15]提出 GRASTA (Grassmannian Robust Adaptive Subspace Tracking Algorithm)以便将背景和前景从随机的次采样数据中分离出来。

1.2 深度学习方法

许多传统方法虽然在运动目标检测时能够取得较好效果,但大多只适用于一些特定场景。随着深度学习的兴起,越来越多的学者开始尝试利用人工神经网络进行运动目标检测。Dosovitskiy 等^[16]设计了名为“FlowNet”的深度网络。该网络能够实现光流的端到端计算,且其精度要高于大部分传统方法。基于 FlowNet 的思想,Ranjan 等^[17]将神经网络与空间金字塔结合起来以检测运动过快的目标;Ahmadi 等^[18]则基于特定的损失函数设计了一个无监督的光流计算网络。

2 本文运动目标检测方法

虽然研究人员已经对利用神经网络计算光流进行了探索,但并未构建真正的运动目标检测网络。本文设计了一个端到端的运动目标检测网络 MONet 以检测间歇性运动的目标。

2.1 网络结构

MONet 的网络结构如图 1 所示。该网络以视频中的相邻两帧作为输入,主要包含编码器与解码器两部分。编码器包括 6 层卷积步长为 2 的卷积层。除第一层采用 5×5 卷积核外,其余各层均采用 3×3

的卷积核。每个卷积层以 Relu 函数为激活函数。卷积部分能够大面积地聚集输入图像的信息, 以寻找到像素点之间的运动关系。但是, 随着卷积层的加深, 网络的分辨率会逐渐降低。因此, 网络还包含一个解码器。该部分用来解码卷积部分得到的高

级特征。此外, 为了提高网络的预测精度, 我们将每层预测结果的上采样、该层的反卷积结果以及对卷积部分的结果共同作为下一反卷积层的输入。通过这种方式, 网络融合了低分辨率特征图的概括性信息和高分辨率特征图的细节性信息。

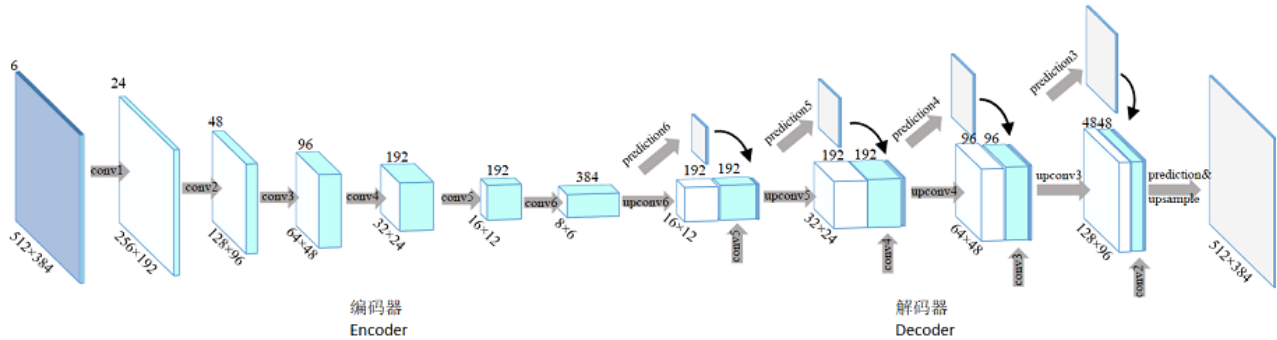


图 1 MONet 网络结构
Fig. 1 MONet network structure

2.2 数据集

与传统方法不同, 神经网络需要依靠大量数据进行训练。为了提供足够的训练数据, 本文合成成了一个简单的数据集-GoChairs。其中, 背景是从网络上收集的 1 800 张图片, 共分为 6 种场景: 海滨、原野、街道、大山、高楼和丛林。每种场景包含 300 张分辨率为 512x384 的图片。同时, 我们对每个场景进行旋转变换, 共产生 3 600 个背景。随后, 我们从椅子模型数据集 renderedchairs^[19]中选取 420 把椅子作为待检测的运动目标。首先, 我们随机选择 0~2 把椅子作为目标, 并将其添加到背景图片上。为了生成运动, 我们利用随机采样得到的参数对椅子进行二维平移变换, 并将其添加到背景图片上。我们将每个背景复用 4 次, 共生成 14 400 个样本。图 2 是 GoChairs 数据集中的部分样本。其中, 每行代表一个测试样本, 前 2 列分别表示第 1 帧和第 2 帧, 第 3 列是运动目标的真实值。

训练集和测试集两个部分。其中, 训练集包含 11 520 个样本, 测试集包含 2 280 个样本。对于 MONet, 我们使用修改后的 caffe 框架进行搭建, 并在 Titan X 平台上完成训练。网络训练时, Adam 被选作优化方法, 基础学习率为 1e-4, batch_size 设置为 8, 最大训练次数设置为 7 200 次。通过实验, MONet 的训练耗时为 10'45''。

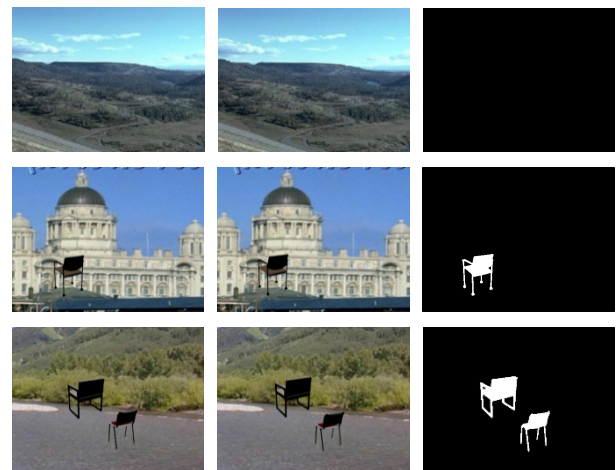


图 2 GoChairs 数据集
Fig. 2 GoChairs dataset

3 实验

3.1 网络训练

我们按照 4:1 的比例将 GoChairs 数据集分成

3.2 测试

首先, 我们利用 GoChairs 中的测试样本对

MONet 的性能进行评估。其后，为了考察网络的泛化性，我们从运动目标检测数据集 CDnet^[20]和 I2R^[21]中选取部分样本对其进行测试。进一步地，针对运动目标检测中存在的间歇性运动问题，我们从定性和定量两个角度评估 MONet 的性能。

3.2.1 GoChairs 测试

图 3 是 MONet 在 GoChairs 中部分测试样本上的测试结果。结果表明，该网络能够根据两幅图像之间像素点的对应关系准确地区分背景与运动目标。当运动目标颜色较浅或与背景颜色值相接近时，该网络检测出的运动目标虽然存在部分缺失，但这并不影响网络的总体特性。需要注意的是，测试集中的背景与训练集中的背景相同。

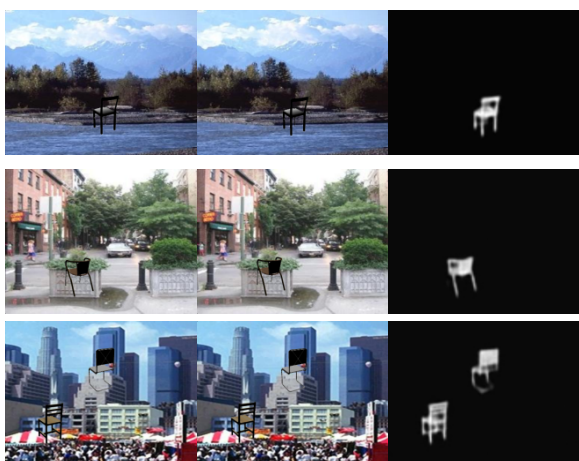


图 3 GoChairs 数据及测试结果
Fig. 3 GoChairs dataset and test results

3.2.2 泛化性测试

由于 MONet 基于 GoChairs 进行训练，所以在 GoChairs 测试集上的测试结果并不足以客观地体现其性能。为此，我们在 CDnet 和 I2R 上进行实验以验证该网络的泛化性能，即 MONet 检测其他类型背景中非椅子类型的运动目标的能力。其中，CDnet 由 31 个真实世界的视频组成，涵盖 6 个类别；I2R 包括九个视频序列，涵盖多种场景。具体地，我们选择 pedestrians, PETS2006, boats 和 curtain 4 个样本对网络进行测试。4 个样本的测试结果如图 4 所示。

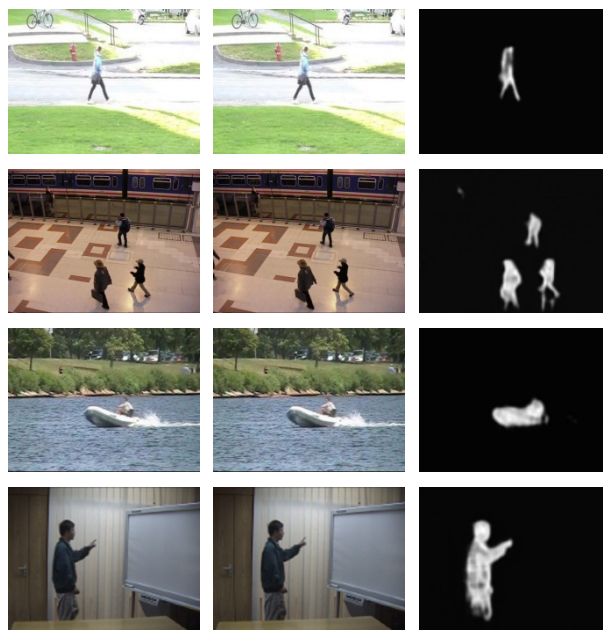


图 4 MONet 泛化性测试
Fig. 4 MONet generalization testing

由图 4 可以看出，网络在 CDnet 和 I2R 上依然取得了很好的效果。这表明：1) 尽管网络只在 GoChairs 上进行了训练，但其具有良好的泛化性；2) 在缺乏训练数据的情况下，利用合成数据集对网络进行训练是十分有效的手段。

3.2.3 间歇性运动测试

针对间歇性运动，从 CDnet 中选取 abandedbox, streetlight 和 winterdriveway 3 个样本对网络进行测试。同时，我们定性和定量地比较 MONet 与其他经典方法在解决该问题上的性能差异。定性分析依赖于比较各种算法的可视化结果，定量分析则依赖于计算各种算法检测结果的 F-measure。F-measure 定义如下：

$$F = 2 \frac{R \cdot P}{R + P} \quad (1)$$

R 和 P 分别指召回率和精度，定义如下：

$$R = \frac{TP}{TP + FN}, P = \frac{TP}{TP + FP} \quad (2)$$

式中： TP 为恢复的真实前景像素的数量； FN 为丢失的真实前景像素的数量； FP 为错误地检测为前景的像素的数量。

我们将 MONet 与 GMM^[7], DECOLOR^[22],

RPCA^[13] 3 种经典算法进行比较。图 5~7 分别为上述方法在 *abandonedbox*、*tramstop* 和 *winterdriveway* 3 个样本上的测试结果。

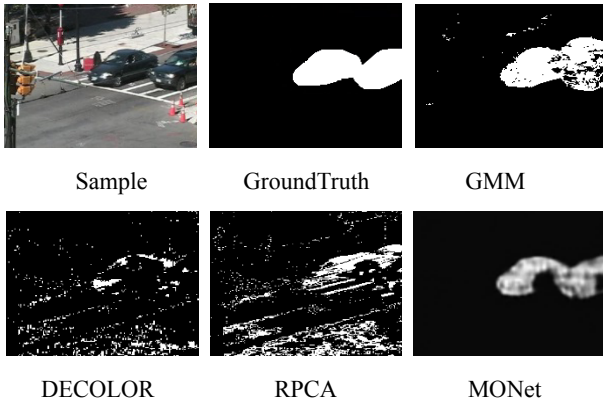


图 5 Abandonedbox 测试结果
Fig. 5 Abandonedbox test results

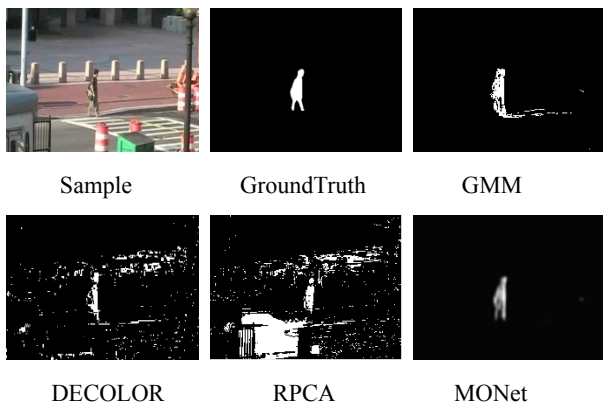


图 6 Tramstop 测试结果
Fig. 6 Tramstop test results

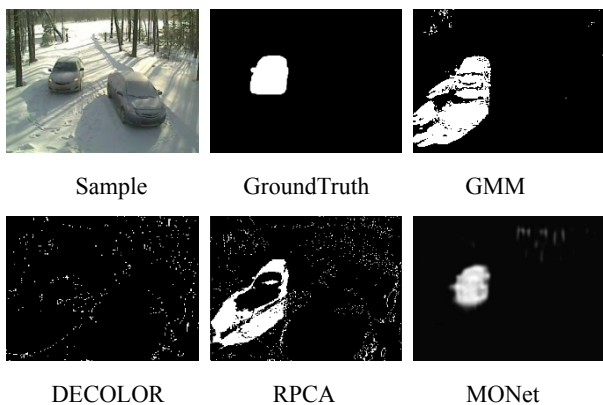


图 7 Winterdriveway 测试结果
Fig. 7 Winterdriveway test results

结果表明, 与传统的方法相比, MONet 在处理目标的间歇性运动问题上更加有效。其原因在于传统方法均根据前几帧图像建立背景模型, 所以当目标由背景状态突然转入运动状态时, 传统方法并不能及时地进行背景的更新。具体地, DECOLOR 算法始终将运动目标误判为背景的一部分。GMM 虽然能够进行一定程度上的背景更新, 但效率很低。相比之下, RPCA 虽然根据整个视频建立背景模型, 但由于背景矩阵的秩并不为 1, 所以其效果也较差。与上述方法不同, MONet 只根据前后两帧图像之间像素点的跃迁关系预测运动目标, 因而能够取得更好的效果。

表 1 是 4 种方法在 3 个测试测试样本上的得分 F-measure。由表可以看出, MONet 在预测准确性上远高于其他算法。GMM、DEXCOLOR 和 RPCA 虽然在静态背景下已经取得了较好效果, 但在间歇性运动问题上, 其得分均低于 0.6。相反, MONet 的得分均高于 0.8。其在 *tramstop* 样本上甚至取得了 0.93 的综合得分。该结果从定量的角度上进一步证明了 MONet 结构的有效性。此外, 其他 3 种算法在预测运动目标时需要花费大量时间, 而 MONet 则具有高度的实时性。

表 1 各种算法的得分
Tab. 1 Scores of various algorithms

Method	Abandonedbox	Tramstop	Winterdriveway
GMM	0.75	0.59	0.41
DECOLOR	0.16	0.28	0.04
RPCA	0.28	0.04	0.21
MONet	0.89	0.93	0.88

4 结论

本文针对运动目标检测问题构建了一个深度卷积神经网络 MONet, 并人工合成了一个数据集 GoChairs 以便进行网络训练。在 GoChairs 数据集、CDnet 数据集和 I2R 数据集上的测试结果表明:

1) 在缺乏训练数据的情况下, 利用合成的数据集对网络进行训练是有效的;

2) 本文构建的深度卷积神经网络 MONet 具

有良好的泛化性;

3) 与传统方法相比, MONet 能够更有效地检测出做间歇性运动的目标。然而, 该网络在某些情况下存在前景缺失的情况。在未来的工作中, 尝试在原有网络上增加补全结构是研究的重点方向。

参考文献:

- [1] Pankaj D. Real-Time Surveillance for Critical Activity Detection in ICUs[C]. Proceedings of the Second International Conference on Computer and Communication Technologies. New Delhi: Springer, 2016.
- [2] Tian B, Morris B T, Tang M, et al. Hierarchical and Networked Vehicle Surveillance in ITS: A Survey[J]. IEEE Transactions on Intelligent Transportation Systems (S1558-0016), 2017, 18(1): 25-48.
- [3] Robertson N, Reid I. A general method for human activity recognition in video[J]. Computer Vision and Image Understanding (S1077-3142), 2006, 104(2/3): 232-248.
- [4] Nait-Charif H, Mc Kenna S J. Head Tracking and Action Recognition in a Smart Meeting Room[J]. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (S2157-491X), 2003.
- [5] C H Anderson, P J Burt, G S van der Wal. Change detection and tracking using pyramid transform techniques[C]. Cambridge: Proc. SPIE, 1985.
- [6] Wren C, Azarbayejani A, Darrell T, et al. Pfunder: real-time tracking of the human body[J]. Transaction on Pattern Analysis and Machine Intelligence (S1939-3539), 1996.
- [7] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking[C]. Proc Computer Vision & Pattern Recognition, IEEE, 1999.
- [8] Elgammal A, Harwood D, Davis L. Non-parametric Model for Background Subtraction[C]. European Conference on Computer Vision. Berlin: Springer, 2000.
- [9] Barnich O, Droogenbroeck M V. VIBE: A powerful random technique to estimate the background in video sequences[C]. International Conference on Acoustics, Speech and Signal Processing. Taipei: IEEE, 2009.
- [10] Candes E, Li X, Ma Y, et al. Robust principal component analysis[J]. Journal of the ACM (S1557-735X), 2011, 58(3): 11.
- [11] Ding X. Bayesian Robust Principal Component Analysis [J]. IEEE Transactions on Image Processing (S1941-0042), 2011, 20(12): 3419.
- [12] Jun He. Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video[C]. CVPR. Providence: IEEE, 2012, 157: 1568-1575.
- [13] Horn B K P, Schunck B G. Determining Optical Flow[C]. Artificial Intelligence, Washington, 1981.
- [14] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]. International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 1981.
- [15] Bouguet J Y. Pyramidal implementation of the affine Lucas-Kanade feature tracker description of the algorithm[R]. Intel Corporation, 2001.
- [16] Dosovitskiy A. FlowNet: Learning optical flow with convolutional networks[C]. ICCV, 2015.
- [17] Ranjan A, Black M J. Optical Flow Estimation using a Spatial Pyramid Network[C]. CVPR, 2016.
- [18] Ahmadi A, Patras I. Unsupervised convolutional neural networks for motion estimation[C]. ICIP. Phoenix: IEEE, 2016.
- [19] Aubry M. Seeing 3D Chairs: Exemplar Part-Based 2D-3D Alignment Using a Large Dataset of CAD Models [C]. CVPR. Columbus: IEEE, 2014.
- [20] Goyette N, Jodoin P M, Porikli F, et al. Changedetection.net: A new change detection benchmark dataset[C]. CVPR. Providence: IEEE, 2012.
- [21] Li L. Statistical modeling of complex backgrounds for foreground object detection[J]. IEEE Transactions on Image Processing (S1941-0042), 2004, 13(11): 1459.
- [22] Zhou X, Yang C, Yu W. Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation[J]. Transactions on Pattern Analysis and Machine Intelligence (S1939-3539), 2013, 35(3): 597-610.