

基于深度自编码网络的运动目标检测

徐培*, 蔡小路, 何文伟, 谢易道

(电子科技大学 计算机科学与工程学院, 成都 611731)

(*通信作者电子邮箱 xupeit268@sina.com)

摘要:针对从动态背景中提取前景效果较差的问题,提出了一种基于深度自编码网络的运动目标检测方法。首先,用一个三层的深度自编码网络从视频图像中提取不包含运动目标的背景图像,将背景图像作为变量构造了深度自编码网络的代价函数;然后,构造了一个分离函数得到了输入图像的背景图像,再用另一个三层的深度自编码网络学习提取出的背景图像;为了使深度自编码网络的学习能够在线地提取运动目标,还提出了一种在线学习算法,通过寻找对代价函数敏感度较低的权重进行合并,从而能够对更多的视频图像进行处理。实验结果表明,所提方法在从动态背景中提取出前景运动目标上相比Lu等的前景检测的工作(LU C, SHI J, JIA J. Online robust dictionary learning. Proceeding of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2013:415–422)检测的准确率提高了6%,并且误报率降低了4.5%。在实际的应用中,能够获得更好的前景背景分离效果,为视频分析等方面的研究奠定更好的基础。

关键词:运动目标检测;视频图像;深度自编码网络;在线学习;代价函数敏感度

中图分类号: TP391.413 **文献标志码:** A

Motion detection based on deep auto-encoder networks

XU Pei*, CAI Xiaolu, HE Wenwei, XIE Yidao

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan 611731, China)

Abstract: To address the poor results of foreground extraction from dynamic background, a motion detection method based on deep auto-encoder networks was proposed. Firstly, background images without containing motion objects were subtracted from video frames using a three-layer deep auto-encoder network whose cost function contained background as variable. Then, another three-layer deep auto-encoder network was used to learn the subtracted background images which are obtained by constructed separating function. To achieve online motion detection through deep auto-encoder learning, an online learning method of deep auto-encoder network was also proposed. The weights of network were merged according to the sensitivity of cost function to process more video frames. From the experimental results, the proposed method obtains better motion detection accuracy by 6%, and lower false rate by 4.5% than Lu's work (LU C, SHI J, JIA J. Online robust dictionary learning. Proceeding of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE Press, 2013:415–422). This work also obtains better extraction results of background and foreground in real applications, and lays better basis for video analysis.

Key words: motion detection; video image; deep auto-encoder network; online learning; sensitivity of cost function

近年来,随着监控视频的普及,视频监控系统越来越多地应用到安全系统当中(如防火、防盗等),并发挥着越来越重要的作用。运动目标(即,前景)检测是视频监控基本的预处理步骤之一,通常是从已有的背景图像中分离出来的。在前人已有的工作中^[1-6],背景图像的像素值往往都假定在固定区间内变化的值,这些关于背景建模的工作包括:混合高斯(Mixture of Gaussian, MOG)^[1]、核方法(Kernel Method)^[2]、区块相关(Block Correlation)^[3]、码书模型(Codebook Model)^[4]、隐马尔可夫模型(Hidden Markov Model)^[5]、线性自回归(Linear Auto Regression)^[6-7]等。虽然这些工作在某些应用场景中取得了较好的效果,但是,这些工作均是基于静态背景的运动目标检测,一旦在背景中出现有周期重复运动的情况,这些方法往往不能取得好的效果。

最近,深度学习(Deep Learning)^[8-10]在学术上引起了广泛的关注。深度模型在目标检测^[11-12]和识别领域^[13]已经展现出了对高维输入数据较强的学习能力。在运动目标检测上,Taylor等^[14]提出用条件受限的玻尔兹曼机(Conditional Restricted Boltzmann Machine, CRBM)来学习行人的运动;在文献[15]中,提出了一种基于弱监督的受限玻尔兹曼机(Weakly Supervised RBM)。这两个工作均是采用学习前景的方式来实现运动目标检测的,但是,在视频中,运动的目标千差万别,如果为每一种运动目标都训练一个网络,显然是不实用的。在文献[16-19]中,运动目标是通过学习背景模型来得到的,在这些工作中,首先用背景图像来学习一个线性相关的字典;然后再基于这些字典所建模的背景提取出运动目标。但是,在实际应用中,干净的背景图像(即:没有前景的背景

收稿日期:2014-05-07;修回日期:2014-06-16。 基金项目:中央基本业务经费资助项目(ZYGX2012YB028)。

作者简介:徐培(1986-),男,四川自贡人,博士研究生,主要研究方向:计算机视觉、机器学习; 蔡小路(1990-),男,湖北黄冈人,硕士研究生,主要研究方向:计算机视觉、机器学习; 何文伟(1988-),男,四川泸州人,硕士研究生,主要研究方向:计算机视觉、机器学习; 谢易道(1988-),男,四川成都人,硕士研究生,主要研究方向:计算机视觉、机器学习。

图像)往往是无法得到的,例如在交通拥挤的十字路口。

针对上述问题,本文提出了一种基于深度自编码网络(Deep Auto-encoder Network)的运动目标检测方法,首先用一个三层的深度自编码网络(背景提取网络)从带有运动目标的图像中提取出干净的背景,然后将这些干净的背景用另一个三层的深度自编码网络(背景学习网络)进行训练来学习背景。为了对视频图像进行在线处理,还提出了基于深度自编码网络的在线学习方法,通过在已训练好的网络中修改对代价函数敏感度较低的权重来提升在线学习的速度。

1 深度自编码网络

本文主要采用文献[8]中所提出的深度自编码网络作为基本结构。对于一个深度自编码网络,对数据的处理主要分为两个阶段:编码和解码:在编码阶段,输入数据 $x \in [0,1]^N$ 通过编码函数 $h = f(x)$ ($h \in [0,1]^M$),其中,编码函数定义为:

$$h = f(x) = \text{sigm}(Wx + b) \quad (1)$$

其中: $W \in \mathbf{R}^{M \times N}$ 代表权重矩阵, $b \in \mathbf{R}^M$ 代表隐层的偏置向量, $\text{sigm}(z) = 1/(1 + \exp(-z))$ 表示 sigmoid 函数;在解码阶段,网络的重构输入 $\hat{x} \in [0,1]^N$ 是通过解码函数(式(2))进行计算得到的。

$$\hat{x} = g(h) = \text{sigm}(W^T h + c) \quad (2)$$

其中: $c \in \mathbf{R}^N$ 代表偏置向量。这里, W, b, c 代表网络的参数,这些参数是通过优化网络的代价函数

$$E(x) = - \sum_{i=1}^N (x_i \ln \hat{x}_i + (1 - x_i) \ln(1 - \hat{x}_i)) \quad (3)$$

来进行训练的,式(3)中的代价函数称为交叉熵函数^[9]。

2 运动目标检测

本文所提出的用深度自编码网络进行运动目标检测的网络结构如图1所示。首先每个视频图像转换为一维向量,并将像素值从0~255转换到[0,1]区间,如果是RGB图像则转换为灰度图像。多个视频图像用集合表示为: $x = \{x^1, x^2, \dots, x^D\}$ ($x^j \in [0,1]^N$),其中: D 表示图像的数目, N 表示向量的维数。 x 作为背景提取网络的输入, $B^0 \in [0,1]^N$ 代表提取出来的背景图像。 H_1 代表编码之后的输出, \hat{x} 代表解码的输出,也称作对输入 x 的重构。在图1中,分离函数 $S(\hat{x}_k^j, B_k^0)$ 用于从重构输出 \hat{x} 中分离出背景学习网络的输入,其中 $k = 1, 2, \dots, N$ 。 H_2 和 \hat{B} 分别表示背景学习网络的编码输出和解码输出。

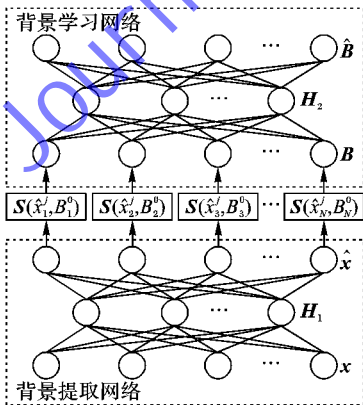


图1 运动目标检测的深度自编码网络结构

2.1 背景提取

在有运动目标的视频图像中,对于相对静止的背景来说,运动的前景目标总是短暂出现在图像当中,图像中的像素在

绝大多数的时间都处于背景的状态当中。基于这样的假设,背景提取网络的代价函数可以定义为:

$$L(x^j, \theta_1) = E(x^j) + \frac{1}{2} \sum_{i=1}^N \|\hat{x}_i^j - B_i^0\|^2 \quad (4)$$

其中: \hat{x}^j 是第 j ($j = 1, 2, \dots, D$) 个输入向量对应的重构输出; D 是输入网络进行训练的向量个数,也即是输入网络训练的視頻图像数目。式(4)中的第二部分称为背景的构造误差,用重构出来的图像来计算背景图像 B^0 ,背景提取网络中的参数是: $\theta_1 = \{W_1, b_1, c_1, B^0\}$,其中 $W_1 \in \mathbf{R}^{M \times N}$, $b_1 \in \mathbf{R}^M$, $c_1 \in \mathbf{R}^N$ 。

进行参数训练时,采用梯度下降算法来进行,即:

$$\theta_1 = \theta_1 - \eta \nabla \theta_1 \quad (5)$$

其中: η 是学习率。 $\nabla \theta_1$ 定义如下:

$$\nabla \theta_1 = \frac{\partial L(x^j, \theta_1)}{\partial \theta_1} = \frac{\partial E(x^j)}{\partial \theta_1} + \frac{\partial \sum_{i=1}^N \|\hat{x}_i^j - B_i^0\|^2}{2 \partial \theta_1} \quad (6)$$

在式(5)、(6)中, θ_1 可以分别替换为 W_1, b_1, c_1, B^0 。当重构误差满足:

$$\sum_{j=1}^D \|\hat{x}^j - x^j\|_2 < \delta \quad (7)$$

成立时,训练结束,其中 $\|\cdot\|_2$ 表示向量的2范数, δ 为设定好的较小的误差阈值。

与传统的深度自编码网络相比,本文所提出的代价函数中包含了背景 B^0 ,但是,向量 B^0 只能表示一个背景图像,在实际应用当中,背景图像的像素并不是绝对不变的,其像素值往往有一些小的波动。为了把背景的变化表达出来,在 B^0 的基础上训练一个背景模型来描述背景的变化,因此,首先定义一个分离函数:

$$B_k^j = S(\hat{x}_k^j, B_k^0) = \begin{cases} \hat{x}_k^j, & |\hat{x}_k^j - B_k^0| \leq \varepsilon_0 \\ B_k^0, & |\hat{x}_k^j - B_k^0| > \varepsilon_0 \end{cases} \quad (8)$$

其中: $k = 1, 2, \dots, N$; $j = 1, 2, \dots, D$, ε_0 是一个阈值参数作为背景的容忍值,这样,对于 D 个输入图像 x^j ($j = 1, 2, \dots, D$),就能够得到 D 个重构输出 \hat{x}^j ,从而得到 D 个背景图像组成的集合 $B = \{B^1, B^2, \dots, B^D\}$ ($B^j \in [0,1]^N$)。相关算法细节见算法1。这里,集合 B 中的图像是从一系列输入图像 x^j 中提取出来的背景,接下来,用另一个深度自编码网络来学习这些提取出来的背景。

2.2 背景学习

从分离函数得到了提取出来的背景图像集 B , B 中的图像作为背景学习网络的输入(如图1所示),为了减少训练时的计算量,背景学习网络的参数 $\theta_2 = \{W_2, b_2, c_2\}$ 用背景提取网络的参数 $\{W_1, b_1, c_1\}$ 来进行初始化。背景学习网络对应的代价函数是:

$$L(B^j, \theta_2) = - \sum_{i=1}^N (B_i^j \ln \hat{B}_i^j + (1 - B_i^j) \ln(1 - \hat{B}_i^j)) \quad (9)$$

其中: $j = 1, 2, \dots, D$; \hat{B}^j 表示背景学习网络的重构输出。

同样地,采用梯度下降算法来训练背景学习网络的参数,即通过公式:

$$\nabla \theta_2 = \frac{\partial L(B^j, \theta_2)}{\partial \theta_2} \quad (10)$$

来计算 $\nabla \theta_2$, θ_2 能够分别被替换成 W_2, b_2, c_2 ,当重构误差满

足: $\sum_{j=1}^D \|\hat{B}^j - B^j\|_2 < \delta$ 时,训练终止。相关细节详见算法1。

算法 1 Background-Learning。

输入: $x = \{x^1, x^2, \dots, x^D\}$ 表示样本图像集, W_1 表示权值矩阵,

B^0 表示背景向量, b_1 和 c_1 分别表示偏置向量。

输出: θ_1, θ_2 。

W_1 随机初始化, B^0 随机初始化, b_1 和 c_1 分别初始化为零向量

Do

For each $j = 1, 2, \dots, D$ do

$H_1 \leftarrow \text{sigm}(W_1 x^j + b_1)$

$\hat{x}^j \leftarrow \text{sigm}(W_1^T H_1 + c_1)$

$\theta_1 \leftarrow \theta_1 - \eta \nabla \theta_1$

End For

While $\left(\sum_{j=1}^D \|\hat{x}^j - x^j\|_2 \geq \delta \right)$

用式(8) 得到 $B = \{B^1, B^2, \dots, B^D\}$

$W_2 \leftarrow W_1, b_2 \leftarrow b_1, c_2 \leftarrow c_1$

Do

For each $j = 1, 2, \dots, D$ do

$H_2 \leftarrow \text{sigm}(W_2 B^j + b_2)$

$\hat{B}^j \leftarrow \text{sigm}(W_2^T H_2 + c_2)$

$\theta_2 \leftarrow \theta_2 - \eta \nabla \theta_2$

end For

While $\left(\sum_{j=1}^D \|\hat{B}^j - B^j\|_2 \geq \delta \right)$

2.3 在线学习

在上文中所介绍的背景提取和背景学习方法中,只能处理 D 帧视频图像,因此,需要执行在线学习的算法,从而在更多的视频图像上进行背景和运动目标的分离。

首先,背景提取网络在在线学习阶段的参数记为 $\theta' = \{W^o, b^o, c^o, B^o\}$,将其分别初始化为 $\theta'^{-h} = \{W^{o-h}, b^{o-h}, c^{o-h}, B^{o-h}\}$,其中 θ'^{-h} 表示已经训练好背景提取网络的参数,在每一轮在线学习中,用 h 帧视频图像来更新背景提取网络。

为了对在线学习进行加速,对权重矩阵 W^o 进行更新。文献[20]中通过计算对代价函数敏感度不同的方式对神经网络进行剪枝,受到该工作的启发,本文的在线学习算法,也通过寻找权重矩阵 W^o 对代价函数的不同敏感度来寻找敏感度较低的权向量。首先, W^o 写为向量的方式 $W^o = [W_1^o, W_2^o, \dots, W_M^o]$,其中 W_j^o 是一个 N 维的向量,定义 ΔL 表示来自权向量 W_j^o 的误差 ΔW_j^o 所引起的代价函数的误差,根据文献[20]的结论,代价函数误差对应的最优化问题是:

$$\min_{\Delta W^o} \Delta L(W^o) = \text{tr} \left(\frac{1}{2} \Delta W^o T H \Delta W^o \right)$$

$$\text{s.t. } \Delta W^o e_j = -W_j^o \quad (11)$$

其中: e_j 表示单位矩阵的第 j 列, H 是赫森矩阵。式(11)的最优解为:

$$\Delta L = W_j^{oT} W_j^o / H_{jj}^{-1} \quad (12)$$

其中: H_{jj}^{-1} 表示赫森矩阵的逆矩阵的第 j 个对角元。 H^{-1} 的计算通过几次迭代处理就能够得到^[20]。

根据式(12),每个权值向量都对应一个 ΔL ,通过不同的 j 对 ΔL 进行排序,对那些满足 $\Delta L < \xi$ 的 W_j^o 随机替换为满足 $\Delta L > \xi$ 的 W_j^o ,其中 ξ 为预设参数。背景学习网络的在线学习算法细节详见算法 2。每次在线学习 h 帧视频图像,图像转换成一维向量之后的大小为 N ,假设迭代次数为 P ,那么在线学习的时间复杂度就为 $O(N \times h \times P)$ 。

算法 2 Online-Learning。

输入: $x = \{x^{t-h+1}, x^{t-h+2}, \dots, x^t\}$ 表示在线输入的样本图像集,

W^{t-h} 表示已训练好的权值矩阵, B^{t-h} 表示已学习出来的背景向量, b^{t-h} 和 c^{t-h} 分别表示已训练好的偏置向量。

输出: W^o, b^o, c^o, B^o 。

$W^o \leftarrow W^{t-h}, b^o \leftarrow b^{t-h}, c^o \leftarrow c^{t-h}, B^o \leftarrow B^{t-h}$

用式(12) 寻找 $\Delta L < \xi$ 的 W_j^o

用满足 $\Delta L > \xi$ 的 W_j^o 随机替换 $\Delta L < \xi$ 的 W_j^o

Do

For each $j = t-h+1, \dots, t$ do

$H^o \leftarrow \text{sigm}(W^o x^j + b^o)$

$\hat{x}^j \leftarrow \text{sigm}(W^{oT} H^o + c^o)$

$\theta^o \leftarrow \theta^o - \eta \nabla \theta^o$

End For

While $\left(\sum_{j=t-h+1}^D \|\hat{x}^j - x^j\|_2 \geq \delta \right)$

2.4 运动目标检测

假设已经训练好了背景提取网络和背景学习网络,对于视频中某一输入图像,按照如下步骤分离其前景和背景,首先,将该输入图像转换成一维向量,记为 y ,分别经过背景提取网络和背景学习网络的计算,得到背景提取网络的输出 \hat{y} 和背景学习网络的输出 \hat{B} ,对应的前景为:

$$F_i = \begin{cases} 0 & |\hat{y}_i - \hat{B}_i| \leq \varepsilon \\ 1 & |\hat{y}_i - \hat{B}_i| > \varepsilon \end{cases} \quad (13)$$

其中: $i = 1, 2, \dots, D$ 表示向量的每一维; ε 是一个预设参数(通常取接近于 0 的正数)。

运动目标检测细节如算法 3 中所示,对于任意一个具有 D' 帧图像的视频,首先调用算法 1 训练两个网络,然后用式(13) 得到前 D 帧图像的前景集 $F = \{F^1, F^2, \dots, F^D\}$,在每一轮的在线学习当中,有 h 帧视频图像调用在线学习算法,然后得到的前景均保存在 F 中。

算法 3 Motion-Detection。

输入: $x = \{x^1, x^2, \dots, x^{D'}\}$ 表示待测视频的所有图像。

输出: F 。

从 x 中选择 D 个图像调用算法 1 训练背景提取网络和背景学习网络。

使用式(13) 得到 D 个图像的前景 $F = \{F^1, F^2, \dots, F^D\}$

$t \leftarrow D + 1$

Do

从 x 取出 h 个样本 $\{x^{t-h+1}, x^{t-h+2}, \dots, x^t\}$

调用算法 2 对 h 个样本进行在线学习

用式(13) 计算 h 个样本的前景 $\{F^{t-h+1}, F^{t-h+2}, \dots, F^t\}$

$F \leftarrow F \cup \{F^{t-h+1}, F^{t-h+2}, \dots, F^t\}$

$t \leftarrow t + h$

End For

While $(t \leq D')$

3 实验

本文采用三个数据集验证相关算法,通过验证算法 3 的结果来和前人的工作进行对比。采用的三个数据集分别是 Ocean^[16]、Rain^[16] 和 Watersurface^[21]。

算法 3 在这三个视频数据集上的实验效果如图 2 所示。图 2 中 (a) 是 Ocean 数据集, (b) 是 Rain 数据集, (c) 是 Watersurface 数据集。从实验效果图中可以看出,本算法能够在视频图像噪声较大的情况下准确地检测出前景,尤其是在 Watersurface 数据集当中,背景里面水面是运动的,但是本文所提出的方法依然能够将行走的行人检测出来。

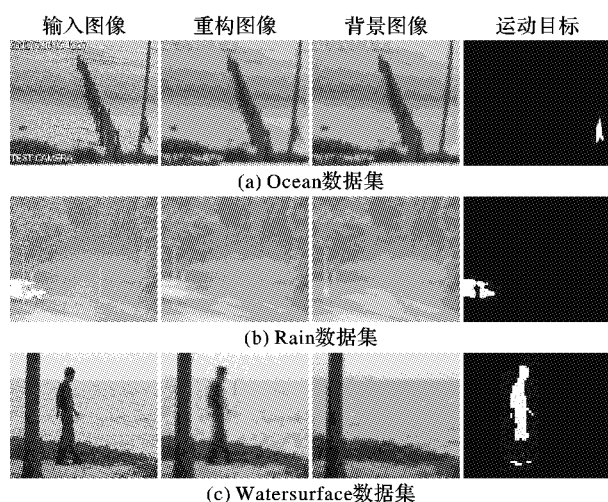


图2 实验效果图

在数据集 Ocean 和 Rain 上采用三个量描述检测效果,分别是 FN(False Negative)、FP(False Positive)、Total(FN + FP),如表 1 和表 2 所示,采用这三个量来和 MOG (Mixture of Gaussian)^[1]、DGS (Dynamic Group Sparsity)^[17]、RDL (Robust Dictionary Learning)^[16] 和 ORDL (Online Robust Dictionary Learning)^[18] 来进行对比。

表1 不同数据集上的对比

数据集	度量指标	MOG	DGS	RDL	ORDL	本方法
Ocean	FN	0.26	0.10	0.73	0.66	0.52
	FP	4.32	1.61	0.21	0.28	0.38
	Total	4.58	1.71	0.94	0.94	0.90
Rain	FN	12.7	4.97	2.91	2.78	2.78
	FP	2.39	2.86	2.31	1.99	1.85
	Total	15.09	4.79	5.04	4.77	4.63

从表 1 中可以看出,在 Ocean 数据集上本方法的 FN 低于 RDL 和 ORDL,但是高于 MOG 和 DGS;本方法的 FP 低于 MOG 和 DGS,高于 RDL 和 ORDL;但是从 FN 和 FP 之和,即 Total 指标上看,本方法是最低的,所以本方法在 Ocean 数据集上是最优的。

同样地,在 Rain 数据集上本方法的 FN 低于 MOG、DGS 和 RDL,且等于 ORDL;本方法的 FP 是最低的;从 FN 和 FP 之和,即 Total 指标上看,本方法是最低的,所以本方法在 Rain 数据集上是最优的。

在 Watersurface 数据集上,本文采用文献[22-23]中所用的度量“F-score”来和前人的工作进行比较,越高的“F-score”说明实验结果越好,“F-score”的定义可参考文献[22-23],之所以采用这个度量是因为前人的相关工作在这个度量上的结果是最好的。

如表 2 所示,在 Watersurface 数据集上,本方法和 Median^[24]、PCP (Principal Component)^[25]、MOG^[1]、DECO (DEtecting Contiguous outliers)^[22] 进行了比较。本方法均高于前人的这些工作。

表3 Watersurface 数据集上的对比

度量指标	Median	PCP	MOG	DECO	本方法
F-score	0.86	0.83	0.47	0.91	0.915

本文所提出的算法,之所以能够比前人的工作得到更好的效果,首先在于本文对背景图像在深度自编码网络上进行

了建模,深度自编码网络具有较强的特征学习能力,在此基础上,通过目标函数学习出背景,为了进一步表达更复杂的背景模型,本文用另一个深度自编码网络来学习变化了的背景,相比前人的工作,具有更强的鲁棒性。这样,使得本文的算法在提取前景时能够更好地排除背景中运动目标的干扰。相关的实验结果也验证了上述结论。

上述实验中,三个数据集的图片大小均为 100×60 ,本算法在 PC 机上每秒钟能够检测出 5 张图片的运动目标。在线学习时,50 张 100×60 大小的图片能够在 20 次迭代之后完成训练,平均耗时 50 s。由此可以看出,本算法在运算速度上能够满足基本的实际应用需要。

4 结语

本文提出了一种基于深度学习的运动目标检测方法,采用两个三层的深度自编码网络对输入视频进行了前景和背景的分隔,为了使深度自编码网络的学习能够在线地提取运动目标,还提出了一种在线学习算法,通过寻找对目标函数敏感度较低的权重进行合并,从而能够对更多的视频图像进行处理。

参考文献:

- [1] STAUFFER C, GRIMSON W E L. Adaptive background mixture models for real-time tracking [C]// Proceeding of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 1999, 2: 246-253.
- [2] MITTAL A, PARAGIOS N. Motion-based background subtraction using adaptive kernel density estimation [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2004, 2: 302-309.
- [3] MATSUYAMA T, OHYA T, HABE H. Background subtraction for non-stationary scenes [C]// Proceeding of the 2000 Asian Conference of Computer Vision. Berlin: Springer-Verlag, 2000: 662-667.
- [4] KIM K, CHALIDABHONGSE T, HARWOOD D, *et al.* Real-time foreground-background segmentation using codebook model [J]. Real-time Imaging, 2005, 11(3): 172-185.
- [5] RITTSCHER J, KATO J, JOGA S, *et al.* A probabilistic background model for tracking [C]// Proceedings of the 2000 European Conference Computer Vision, LNCS 6312. Berlin: Springer-Verlag, 2000: 336-350.
- [6] ZHONG J, SCLAROFF S. Segmenting foreground objects from a dynamic textured background via a robust Kalman filter [C]// Proceedings of the 2003 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2003: 44-50.
- [7] TIAN Y, TIAN S, XU Y, *et al.* Image object detection based on local feature and sparse representation [J]. Journal of Computer Applications, 2013, 33(6): 1670-1673. (田元荣, 田松, 许悦雷, 等. 基于局部特征和稀疏表示的图像目标检测算法[J]. 计算机应用, 2013, 33(6): 1670-1673.)
- [8] BENGIO Y, LAMBLIN P, POPOVICI D, *et al.* Greedy layer-wise training of deep networks [C]// Proceedings of the 20th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2007: 153-160.
- [9] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [10] VINCENT P, LAROCHELLE H, BENGIO Y, *et al.* Extracting and composing robust features with denoising autoencoders [C]// Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008: 1096-1103.

(下转第 2962 页)

- [7] MATHIEU B, MELCHIOR P, OUTSALLOUP A, *et al.* Fractional differentiation for edge detection [J]. *Signal Processing*, 2003, 83(11): 2421–2432.
- [8] ZUO K, SUN T, LI Z, *et al.* 2D fractional Kalman filter and its application to image process [J]. *Journal of Electronics & Information Technology*, 2010, 32(12): 3027–3031. (左凯, 孙同景, 李振华, 等. 二维分数阶卡尔曼滤波及其在图像处理中的应用[J]. *电子与信息学报*, 2010, 32(12): 3027–3031.)
- [9] GAO C, ZHOU J. Image enhancement based on quaternion fractional directional differentiation [J]. *Acta Automatica Sinica*, 2011, 37(2): 150–159. (高朝邦, 周激流. 基于四元数分数阶方向微分的图像增强[J]. *自动化学报*, 2011, 37(2): 150–159.)
- [10] BAI J, FENG X. Fractional-order anisotropic diffusion for image denoising [J]. *IEEE Transactions on Image Processing*, 2007, 16(10): 2492–2502.
- [11] ZHANG J, WEI Z. Fractional-order multi-scale variation PDE model and adaptive algorithm for SAR image denoising [J]. *Journal of Electronics and Information Technology*, 2010, 32(7): 1654–1659. (张军, 韦志辉. SAR 图像去噪的分数阶多尺度变分 PDE 模型及自适应算法[J]. *电子与信息学报*, 2010, 32(7): 1654–1659.)
- [12] ZHANG J, WEI Z. A class of fractional-order multi-scale variational models and alternating projection algorithm for image denoising [J]. *Applied Mathematical Modelling*, 2011, 35(5): 2516–2528.
- [13] CHEN B, WANG W, QIN Q. Infrared dim target detection based on fractional integral operator [J]. *Control and Decision*, 2012, 27(1): 147–151. (陈炳文, 王文伟, 秦前清. 基于分数阶积分算子的红外弱小目标检测[J]. *控制与决策*, 2012, 27(1): 147–151.)
- [14] LIU Y, PU Y, ZHOU J. A digital image denoising method based on fractional calculus [J]. *Journal of Sichuan University: Engineering Science Edition*, 2011, 43(3): 90–95. (刘彦, 蒲亦非, 周激流. 一种基于分数阶积分的数字图像去噪算法[J]. *四川大学学报: 工程科学版*, 2011, 43(3): 90–95.)
- [15] HUANG G, PU Y, CHEN Q, *et al.* Research on image denoising based on fractional order integral [J]. *Systems Engineering and Electronics*, 2011, 33(4): 925–932. (黄果, 蒲亦非, 陈庆利, 等. 基于分数阶积分的图像去噪[J]. *系统工程与电子技术*, 2011, 33(4): 925–932.)
- [16] TAO R, ZHANG F, WANG Y. Progress in discrete fractional Fourier transform [J]. *Science in China Series E: Information Sciences*, 2008, 38(4): 481–503. (陶然, 张峰, 王越. 分数阶 Fourier 变换离散化的研究进展[J]. *中国科学 E 辑: 信息科学*, 2008, 38(4): 481–503.)
- [17] HUANG G, PU Y, CHEN Q, *et al.* Non-integer step of fractional order differential in the image enhancement application [J]. *Journal of Sichuan University: Engineering Science Edition*, 2011, 43(1): 129–136. (黄果, 蒲亦非, 陈庆利, 等. 非整数步长的分数阶微分滤波器在图像增强中的应用[J]. *四川大学学报: 工程科学版*, 2011, 43(1): 129–136.)
- (上接第 2937 页)
- [11] YUAN F. Codebook generation based on self-organizing incremental neural network for image classification [J]. *Journal of Computer Applications*, 2013, 33(7): 1976–1979. (袁飞云. 基于自组织增量神经网络的码书产生方法在图像分类中的应用[J]. *计算机应用*, 2013, 33(7): 1976–1979.)
- [12] OUYANG W, WANG X. Joint deep learning for pedestrian detection [C]// *Proceeding of the 2013 IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2013: 2056–2063.
- [13] LE Q V, ZOU W Y, YEUNG S Y, *et al.* Learning hierarchical invariant spatiotemporal features for action recognition with independent subspace analysis [C]// *Proceeding of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2011: 3361–3368.
- [14] TAYLOR G W, HINTON G E, ROWEIS S T. Modeling human motion using binary latent variables [C]// *Proceedings of the 20th Annual Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2007: 1345–1353.
- [15] HEES N, ROUX N L, WINN J. Weakly supervised learning of background segmentation using masked RBMs [C]// *International Conference on Artificial Neural Networks*, LNCS 6312. Berlin: Springer-Verlag, 2011: 9–16.
- [16] ZHAO C, WANG X, CHAM W K. Background subtraction via robust dictionary learning [EB/OL]. [2014-02-22]. <http://www.docin.com/p-233234564.html>
- [17] HUANG J, HUANG X, METAXAS D N. Learning with dynamic group sparsity [C]// *Proceeding of the 2009 IEEE International Conference of Computer Vision*. Piscataway: IEEE Press, 2009: 64–71.
- [18] LU C, SHI J, JIA J. Online robust dictionary learning [C]// *Proceeding of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway: IEEE Press, 2013: 415–422.
- [19] CEVHER V, SANKARANARAYANAN A, DUARTE M, *et al.* Compressive sensing background subtraction [C]// *Proceedings of the 2008 European Conference on Computer Vision*, LNCS 6312. Berlin: Springer-Verlag, 2008: 155–168.
- [20] XU J, DANIEL W C H. A new training and pruning algorithm based on node dependence and Jacobian rank deficiency [J]. *Neurocomputing*, 2006, 70(1/2/3): 544–558.
- [21] LI L, HUANG W, GU I, *et al.* Statistical modeling of complex backgrounds for foreground object detecting [J]. *IEEE Transactions on Image Processing*, 2004, 13(11): 1459–1472.
- [22] ZHOU X, YANG C, YU W. Moving object detection by detecting contiguous outliers in the low-rank representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(3): 597–610.
- [23] STAUFFER C, GRIMSON W. Adaptive background mixture models for real-time tracking [C]// *Proceeding of the 1999 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 1999: 2246–2252.
- [24] GUTCHESS D, TRAJKOVICS M, COHEN-SOLAL E, *et al.* A background model initialization algorithm for video surveillance [C]// *Proceeding of the 2001 8th IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2001: 733–740.
- [25] CANDES E, LI X, MA Y, *et al.* Robust principal component analysis [EB/OL]. [2014-02-01]. <http://wenku.baidu.com/view/95964f3243323968011c9261.html>