一种分步的融合时空信息的背景建模

储 珺 [1] 杨 樊 [1] 张桂梅 [1] 汪凌峰 [2]

摘 要 自然场景中的光照突变和树枝、水面等不规则运动是背景建模的主要困难. 针对该问题, 提出一种分步的融合时域信息和空域信息的背景建模方法. 在时域, 采用具有光照不变性的颜色空间表征时域信息, 并提出对噪声和光照突变具有较好适应性的码字聚类准则和自适应背景更新策略, 构造了对噪声和光照突变具有较好适应性的时域信息背景模型. 在空域, 通过采样将测试序列图像分成两幅子图, 而后利用时域模型检测其中一幅子图, 并将检测结果作为另一幅子图的先验信息, 同时采用马尔科夫随机场 (Markov random field, MRF) 对其加以约束, 最终检测其状态. 在多个测试视频序列上的实验结果表明, 本文背景模型对于自然场景中的光照突变和不规则运动具有较好的适应性.

关键词 时空背景模型, 前景检测, 马尔科夫随机场, 码本

引用格式 储珺, 杨樊, 张桂梅, 汪凌峰. 一种分步的融合时空信息的背景建模. 自动化学报, 2014, **40**(4): 731−743

DOI 10.3724/SP.J.1004.2014.00731

A Stepwise Background Subtraction by Fusion Spatio-temporal Information

CHU Jun[1] YANG Fan[1] ZHANG Gui-Mei[1] WANG Ling-Feng[2]

Abstract In a natural scene, it is difficult to create a background model for the presence of illumination variation and irregular motions including waving trees, rippling water, etc. This paper proposes a new stepwise algorithm by fusing spatio-temporal information. In the time domain, we characterize the temporal information in the color space which is invariant to photometric changing. On this basis, we propose a clustering criterion of codeword which is adaptive to noise and illumination variation, and present a novel adaptive background updating strategy. Then a temporal information background model which has a better adaptability to noise and photometric invariants is constructed. In the spatial domain, we first divide the test frame into two sub-images by sampling and then utilize temporal information to detect one of them. Furthermore, we regard the detection results as priori information of the other sub-image and adopt Markov random field to restrict it simultaneously, then detect its state. Extensive experiments are conducted on several test video sequences. Compared with the mixture of Gaussians (MOG), standard codebook model (SCBM), and improved codebook model (ICBM), the results show that out algorithm has better adaptability to the illumination variation and irregular movement in natural scenes.

Key words Spatio-temporal background model, foreground detection, Markov random field (MRF), codebook

Citation Chu Jun, Yang Fan, Zhang Gui-Mei, Wang Ling-Feng. A stepwise background subtraction by fusion spatio-temporal information. *Acta Automatica Sinica*, 2014, **40**(4): 731−743

从视频序列中检测运动目标是智能视频监控[1]、基于视觉的人机交互以及人体姿态分析[2] 等领域的重要研究内容. 目前背景剪除法是常用的运动目标检测方法之一. 背景剪除法通过计算当前帧与背景模型的差异来区分背景和运动目标, 其性能主要依赖于所采用的背景建模技术. 目前的背景建模主要有基于时域信息和时空域信息融合的方法, 它们能适应场景中的缓慢光照变化和规则的运动, 但对自然场景中光照的突然改变、树枝的不规则摇摆和水面的闪烁等非平稳变化, 还没有很好的解决方案.

时域信息反映了场景在过去时间里的统计特性, 根据这种统计特性可以描述场景短期未来的状态. 如 Wren 等[3] 和Horprasert 等[4] 提出采用单高斯背景模型来描述像素观测值, 该方法对照度的变化和视频中的噪声有较好的适应性. 但是被监测场景的变化通常是多模态的, 因此, Stauffer 等[5] 提出采用混合高斯模型 (Mixture of Gaussians, MOG) 描述背景的多模态变化, 并且采用固定的学习率更新背景. 但是该方法中学习率的选取是一个关键问题,

收稿日期 2012-08-30 录用日期 2012-12-24
Manuscript received August 30, 2012; accepted December 24, 2012
国家重点基础研究发展计划 (973 计划) (2009CB320902), 国家自然科学基金 (61263046), 中国航天科技集团公司航天科技创新基金 (CASC201102) 资助
Supported by National Basic Research Program of China (973 Program) (2009CB320902), National Natural Science Foundation of China (61263046), and the Aerospace Science and Technology Innovation Fund of China (CASC201102)
本文责任编委 封举富
Recommended by Associate Editor FENG Ju-Fu
1. 南昌航空大学计算机视觉研究所 南昌 330063 2. 中国科学院自动化研究所 北京 100190
1. Institute of Computer Vision, Nanchang Hangkong University, Nanchang 330063 2. Institute of Automation, Chinese Academy of Sciences

如果模型更新速度过快, 运动较慢的物体会被误检测为背景; 更新速度过慢, 则刻画背景的模型自适应能力较弱. 针对该问题, Chen 等[6] 提出对背景和前景采用不同的学习率进行更新. 此外, Kim 等[7] 将背景中像素每一模态对应的观测值、亮度、出现频率以及访问时间等参数表征为码本中的一个码字, 通过计算颜色失真度以及亮度波动范围来聚类码字, 建立了码本模型 (Codebook model). 以上几种背景模型在场景光照出现突然变化时, 往往误检测率较高, 其原因在于: 1) 表征背景时域信息的特征不具有光照不变性; 2) 没有采用有效的背景模型更新方法, 使得变化的背景不能立即融入背景模型. 赵旭东等[8] 提出一种适应户外光照变化的背景建模方法, 该方法建立了基于帧间像素亮度差统计直方图的像素亮度扰动阈值, 提出基于自回归的背景模型参数更新方法, 增加了背景模型适应户外光照变化的能力.

监测场景中不规则的运动是室外场景背景建模技术的另一个难点. 通过融合时空信息来建立背景模型是一种较好的解决策略. 如 Yin 等[9] 提出的背景模型将以当前帧为中心的前后三帧的时域信息和当前帧的空域信息的相关关系统一到一个马尔科夫随机场 (Markov random field, MRF) 下, 建立了三维的背景模型. Migdal 等[10] 构建了三个马尔科夫随机场模型, 分别用来描述背景当前帧和前一帧的时域信息、当前帧和后一帧的时域信息及当前帧的空域信息. Wu 等[11] 通过马尔科夫随机场, 建立了一个融合背景当前帧以及前 4 帧的时域信息和当前帧以及前一帧的空域信息的背景模型.

以上几种背景模型都是通过马尔科夫随机场融合了相邻帧的时域信息和空域信息. 但实际上, 当背景中存在不规则的运动时 (如树叶的随风摆动、水面的闪烁等), 相邻帧和当前帧的状态可能不再具有很高的相关性, 此时再利用相邻帧的信息作为先验信息很可能会带来较高的误检测率, 但是在同一帧内, 相邻像素之间存在较高的相关性. 为了较好地利用同一帧内相邻像素之间的相关性, 本文提出了一种分步的、融合当前帧背景的时域信息和空域信息的背景建模方法. 算法首先提出一种具有光照不变性的时域信息背景模型, 并通过训练图像获取背景的时域模型; 然后, 通过采样将测试视频序列图像分成两幅子图, 利用训练得到的时域模型先检测其中一幅子图, 将检测结果作为另一幅子图的先验信息并采用马尔科夫随机场对其加以约束以检测该幅子图的状态. 本文的算法框图如图 1 所示. 本文的主要贡献如下:

1) 采用具有光照不变性的颜色空间表征时域信息, 构造了具有光照不变性的码字.

2) 构建了关于码字亮度和色度分量的聚类准则. 该准则对图像序列中的噪声以及亮度随场景光照变化的波动具有较好的适应性.

3) 提出一种新的背景更新策略. 利用回归分析, 根据当前帧的像素值和与之匹配的码字的变化差异, 实现动态背景的更新.

4) 提出一种分步的、融合背景时空信息的自然场景背景建模方法. 通过采样将测试序列图像分成两幅子图, 采用分步的策略, 先利用训练得到的时域模型检测其中一幅子图, 并将检测结果作为另一幅子图的先验信息, 同时采用马尔科夫随机场约束该先验信息, 以检测该幅子图状态. 利用了当前帧空间

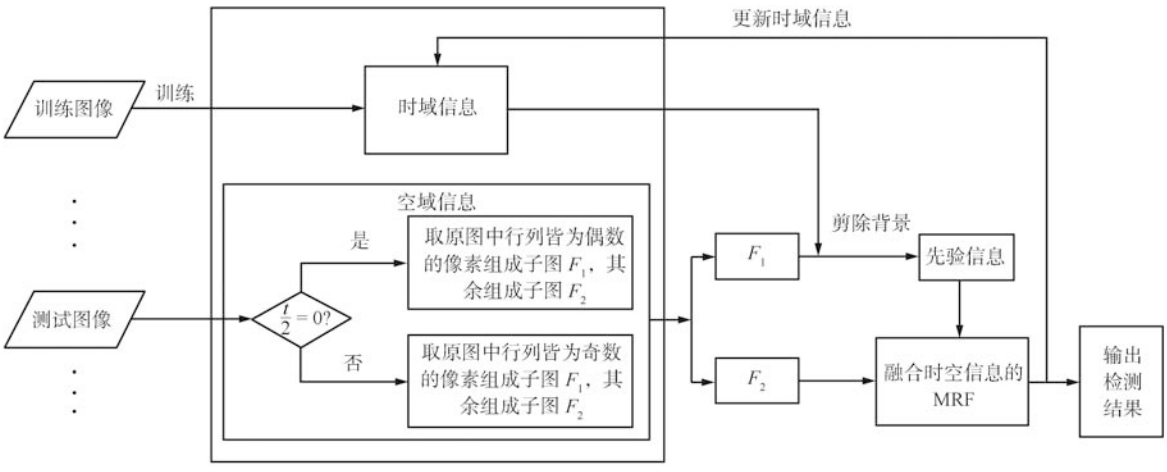


图 1 融合时空信息的背景模型算法框图

Fig. 1 Flowchart of background subtraction by fusion spatio-temporal information

的高度相关性, 同时避免了相邻帧空间的低相关性影响.

1　基于光照不变码本的时域背景模型

1.1　光照不变码本

在 Kim 等[7] 提出的码本模型中, 将图像序列每一像素表征为一个码本 $C=\{c_1,c_2,\cdots,c_L\}$, 该像素时域信息中每一种变化模态在 RGB 空间中的观测值、亮度、出现频率以及访问时间等参数表征为码本中的一个码字 c_j $(j=1,\cdots,L,\ L$ 为码字的个数)[1]. 但是通常 RGB 空间对光照变化敏感, 从而导致在 RGB 空间中的码字聚类使得该模型对光照变化适应性较差.

为了解决场景中光照突变的问题, 本文引入了一种具有光照不变性的颜色空间 (I,E,W)[12] . 该颜色空间和 (R,G,B) 颜色空间的转化关系如下:

$$\begin{bmatrix} I \\ E \\ W \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

其中, I 表示亮度值, E 表示黄蓝通道的色度, W 表示红绿通道的色度, 该转换去掉了 RGB 空间中 R、G、B 三分量的相关性[12].

在该颜色空间下, 我们定义本文的码本为: $C=\{c_1,c_2,\cdots,c_L,F\}$, 其中码本中的码字 c_j (光照不变码字) 定义如下:

$$c_j=\{I_j,E_j,W_j,\hat{\tau}_j,\check{\tau}_j,\sigma.E_j,\sigma.W_j,f_j,\lambda_j,p_j,q_j\} \tag{2}$$

这里 $\hat{\tau}_j(\check{\tau}_j)$ 表示亮度波动范围的上限 (下限), 该阈值是在训练过程中通过自适应学习得到的, $\sigma.E_j$, $\sigma.W_j$ 分别是黄蓝通道和红绿通道的色度分量的方差, f_j 表示该码字出现的频率, λ_j 表示该码字未出现的最长时间间隔, p_j (q_j) 表示该码字第一次 (最后一次) 被访问的时间. 而码本中的元素 F 记录着在该码本中的所有码字的出现频数之和.

1.2　码字聚类准则

传统的码字聚类准则[7] 是计算当前帧的像素值和背景模型中码字的颜色失真度以及亮度的波动范围. 当图像中存在噪声时, 颜色失真度会在较大的范围内波动, 尤其是在照度较低时, 这种波动相对更为明显, 因此, 该方法对噪声的鲁棒性较差. 如果图像中每一像素对应特定光照下背景中的特定平面, 即使在有噪声的情况下, 采用单高斯模型也足以对其在图像序列中得到的观测值建模[5]. 如果将码本的每个码字的色度看作以上特定条件下的观测值, 那么可用正态分布作为当前观测值的色度分量与码字中的色度分量的聚类准则, 即:

$$\boldsymbol{p}(\mu_t^m \mid \mu_{t,j}^m,\Sigma_{t,j}^m)=\frac{1}{(2\pi)^{\frac{d}{2}}\mid\Sigma_{t,j}^m\mid^{\frac{1}{2}}}\times \tag{3}$$
$$\exp\left(-\frac{1}{2}\times((\mu_t^m-\mu_{t,j}^m)^{\mathrm{T}}(\Sigma_{t,j}^m)^{-1}(\mu_t^m-\mu_{t,j}^m))\right)$$

其中, $\boldsymbol{\mu}_t^m=(E_t^m,W_t^m)$ 是第 t 帧图像中第 m 个像素的观察值 $\boldsymbol{f}_t^m=(I_t^m,E_t^m,W_t^m)$ 的色度分量, 该像素对应的码本为 $C_t^m=\{c_{t,1}^m,c_{t,2}^m,\cdots,c_{t,L}^m,F_t^m\}$, $\boldsymbol{\mu}_{t,j}^m=(E_{t,j},W_{t,j})$ 为该码本中码字 $c_{t,j}^m$ 的色度分量, 代表高斯函数的均值, $\Sigma_{t,j}^m$ 为此时高斯函数的协方差, $d=2$ 为 $\boldsymbol{\mu}_t^m$ 的维数.

$$\Sigma_{t,j}^m=\begin{pmatrix} \sigma.E_{t,j} & 0 \\ 0 & \sigma.W_{t,j} \end{pmatrix} \tag{4}$$

$\sigma.E_{t,j}$, $\sigma.W_{t,j}$ 是对应色度分量的方差, 为了计算方便, 在实验中, 判断 $\boldsymbol{\mu}_t^m$ 与码字的 $\boldsymbol{\mu}_{t,j}^m$ 是否满足式 (5):

$$\begin{cases} \mid E_t^m-E_{t,j}\mid\leq\sigma.E_{t,j} \\ \mid W_t^m-W_{t,j}\mid\leq\sigma.W_{t,j} \end{cases} \tag{5}$$

为了应对场景变化程度的差异, 本文通过自适应学习来确定 $\sigma.E_{t,j}$, $\sigma.W_{t,j}$, 首先, 将它们分别初始化为 6.0 和 9.6, 如果当前的 $\boldsymbol{\mu}_t^m$ 与码字的 $\boldsymbol{\mu}_{t,j}^m$ 不满足式 (5), 那么对这两个参数不予更新, 如果满足, 那么 $\sigma.E_{t,j}$, $\sigma.W_{t,j}$ 自适应更新如下:

$$\begin{cases} \hat{\sigma}.E_{t,j}=(1-\rho)\sigma.E_{t,j}+\rho(E_t^m-E_{t,j})^2 \\ \hat{\sigma}.W_{t,j}=(1-\rho)\sigma.W_{t,j}+\rho(W_t^m-W_{t,j})^2 \\ \rho=\dfrac{\alpha F_t^m}{f_{t,j}} \end{cases} \tag{6}$$

其中, $\alpha=0.001$ 是模型的学习率, $\hat{\sigma}.E_{t,j}$, $\hat{\sigma}.W_{t,j}$ 是更新后的方差.

实验发现, 当光照由暗到亮发生变化时, 图像亮度 I 会随着光照的变化出现波动 (如图 2 中的 I 线), 而黄蓝通道和红绿通道的色度 E, W 相对于亮度的变化, 受光照影响较小 (如图 2 中的 E 和 W 线, 为了避免色度的变化曲线与横坐标重合, 将色度的变化曲线进行了适当的平移); 并且随光照变化, 不同颜色物体的亮度变化趋势是一致的, 可以用统一形式的函数来描述; 同时在光照强度较低时, 图像亮度比较小, 且随光照变化的波动范围较小, 在光照强度较强烈时, 图像亮度较大且波动范围较大. 鉴于亮度

[1]在训练阶段, 码字个数由该像素上出现的变化模态个数而定; 在测试阶段, 当更新码本时, 删除了某个已过时 (设定阈值时间范围内未被访问) 的模态, 则码字个数相应减 1, 当创建了新的模态到码本中, 则码字个数相应加 1.

变化的复杂性, 提出一个函数来描述图像序列亮度随场景光照变化的波动范围是比较困难的, 尽管图 2 中三幅图里的亮度变化曲线可以用统一形式的函数来描述, 但是估计它们的参数是一件困难的事.

如图 2 中 I 曲线所示, 当某像素对应码本里码字的亮度因相邻帧间场景中光照变化导致其变化时, 通过设定码字里的亮度在相邻帧间的波动上下限, 可以有效地判断因场景光照变化导致的图像亮度变化. 因此, 本文的亮度聚类准则如下:

$$\text{brightness}(I_t^m, I_{t,j}) = \begin{cases} \text{真}, & \text{若 } \beta_1\check{\tau}_{t,j} \le \tau_t^m \le \beta_2\hat{\tau}_{t,j} \\ \text{假}, & \text{其他} \end{cases} \tag{7}$$

其中, $\tau_t^m = I_t^m - I_{t,j}$, $I_{t,j}$ 为码字 $c_{t,j}^m$ 中的亮度分量, 参数 $\beta_1(\beta_2)$ 用于控制照度变化时引起亮度减弱 (增强) 的适应能力, 在实验中它们的取值分别为 1.5(2.0). 在背景模型训练过程中, 将 $\hat{\tau}_{t,j}(\check{\tau}_{t,j})$ 初始化为 8.0(−8.0). 如果观测值 $\boldsymbol{f}_t^m$ 的色度分量满足式 (5) 且亮度分量在式 (7) 取值为真, 若 τ_t^m 为正数则将其赋给 $\hat{\tau}_{t,j}$, 否则, 赋给 $\check{\tau}_{t,j}$, 以此来更新码字的亮度波动的上下限, 同时该观测值可判为背景, 否则, 为前景.

我们统计了真实场景中 4 个样本点 a、b、c、d 随场景光照变化时的亮度和色度变化情况, 如图 3 所示, 场景光照变化时对亮度的影响较大, 对色度也有影响, 色度受影响后的变化趋势和物体本身的颜色以及表面有关, 但相对于亮度, 色度变化较小. 我们发现只要对色度加以适当更新, 同时对其方差按式 (6) 加以更新, 就能较好应对场景的光照变化.

1.3 背景更新策略

好的时域信息背景模型不仅需要较强的区分前景的能力, 还需要随着背景的变化而动态更新. 当背景中存在照度变化时, 相应区域的色度也会发生些许变化, 虽然这种变化是不确定的, 但是它们之间存在着相关关系. 因此, 本文采用回归分析来描述背景

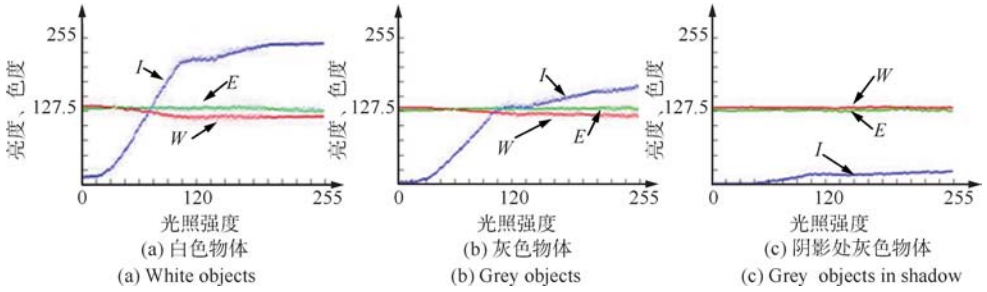


图 2 图像亮度、色度随场景光照变化的关系

Fig. 2 Relationship between brightness and chrominance of image and illumination variation of scene

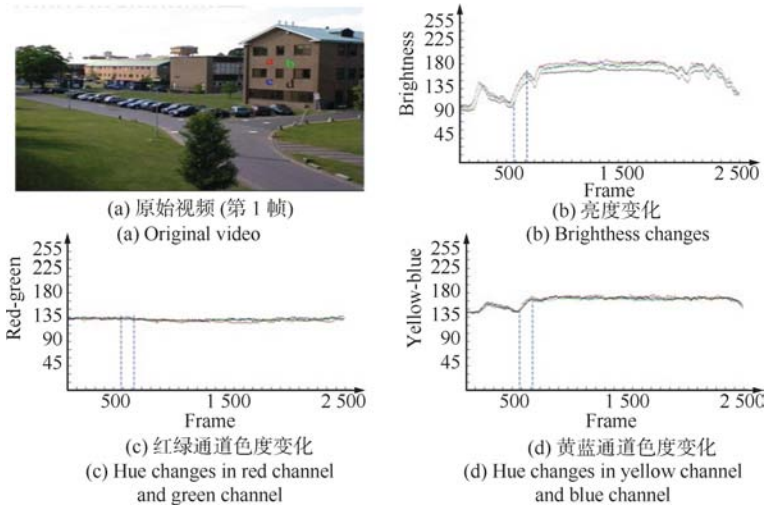


图 3 场景内亮度变化的实验

Fig. 3 Experiment about brightness changes in natural scene

更新, 即根据当前像素值以及和当前像素值匹配的码字的差异来更新该码字.

假设 E_t 是当前观测的黄蓝通道色度值, $\tau_t^E = E_t - E_{t,j}$ 是该观测值和与之匹配的码字 $c_{t,j}$ 对应通道色度值的差异, 更新后该码字的黄蓝通道色度值为 $\hat{E}_{t,j} = f(E_{t,j}, \tau_t^E)$, $\hat{E}_{t,j}$ 是与 $E_{t,j}$, τ_t^E 相关的随机变量.

当背景变化时, 同一物体表面的色度值的变化具有平滑性. 在本文定义的码字聚类准则下, 如果当前观测色度值 E_t 和与之匹配的码字的 $E_{t,j}$ 偏离越大, 则 τ_t^E 相应变化越大, 则 $E_{t,j}$ 需要较大程度的更新才能适应当前场景中的变化; 但是当 E_t 与 $E_{t,j}$ 差异太大时, E_t 属于 $E_{t,j}$ 的隶属度较低, 那么此时需要适当抑制更新程度. 因此, 我们定义黄蓝通道色度更新公式如下:

$$\hat{E}_{t,j} = f(E_{t,j}, \tau_t^E) = a_1 E_{t,j} + b_1 \tau_t^E \left(1 - \exp\left(-\frac{|\tau_t^E|}{\sigma_1^2}\right)\right) \tag{8}$$

从概率论的观点来说, 当 E_t 取任意数时, 它服从正态分布 $\mathrm{N}(\hat{E}_{t,j}, \sigma_4^2)$, 其中, $\hat{E}_{t,j}$ 是该正态分布的均值, 那么随机变量 E_t 的概率密度函数为

$$f(E_t) = \frac{1}{\sqrt{2\pi}\sigma_4} \exp\left(-\frac{1}{2\sigma_4^2}(E_t - \hat{E}_{t,j})^2\right) \tag{9}$$

设 $X = \{(E_{t,j}, E_t, \tau_t^E, \hat{E}_{t,j}) \mid t = 1, \cdots, n\}$ 为某像素对应码本中与之匹配的第 j 个码字的训练样本集, 假如用于训练背景模型的图像序列有 100 帧, 而属于对应像素码本中第 j 个码字的样本只有 20 个, 那么训练样本集 X 中的 n 为 20, 样本集 X 中的 $E_{t,j}$ 为判断第 t 个观测样本时码字中的黄蓝通道色度, 此时, E_t 为第 t 个观测样本, $\hat{E}_{t,j}$ 为按式 (8) 更新后的值, 作为更新下一个与之紧邻的观测样本时码字中的黄蓝通道色度值. 将该样本集代入上述概率密度函数, 其似然函数为

$$L(a_1, b_1) = \prod_{t=2}^{n} f(E_t) = \left(\frac{1}{\sqrt{2\pi}\sigma_4}\right)^{n-1} \times \exp\left(-\frac{1}{2}\sum_{t=2}^{n}\left(\frac{(E_t - \hat{E}_{t,j})^2}{\sigma_4^2}\right)\right) \tag{10}$$

要使似然函数 $L(a_1, b_1)$ 取得最大值, 需使得:

$$S(a_1, b_1) = \sum_{t=2}^{n}\left(\frac{(E_t - \hat{E}_{t,j})^2}{\sigma_4^2}\right)$$

最小, 为了计算简便, 这里参数 σ_1, σ_4 视为常数, 则 a_1, b_1 的计算过程如下:

步骤 1. 初始化 a_1, b_1, σ_1.

步骤 2. 将 a_1, b_1 带入目标函数 $S(a_1, b_1)$, 得 $S_1 = R_1 = S(a_1, b_1)$, 固定 b_1, 令 $a_1.p = a_1$, $a_1 = a_1.p - \delta$, 得 $R_2 = S(a_1, b_1)$, 若 $R_2 - R_1 > 0$, 则 $a_1 = a_1.p + \delta$, 固定 a_1, 得 $R_1 = S(a_1, b_1)$, 令 $b_1.p = b_1$,$b_1 = b_1.p - \delta$, 得 $R_2 = S(a_1, b_1)$, 若 $R_2 - R_1 > 0$, 则 $b_1 = b_1.p + \delta$, 计算 $S_2 = S(a_1, b_1)$.

步骤 3. 若 $|S_1 - S_2| > \gamma$, 则转步骤 2, 否则结束, 输出 a_1, b_1.

步骤中参数 $\delta = 0.001$ 为一微小增, $\gamma = 0.01$ 为判断优化步骤是否结束的阈值. 同理可得红绿通道色度和亮度的更新公式为

$$\begin{cases} \hat{W}_{t,j} = f(W_{t,j}, \tau_t^W) = a_2 W_{t,j} + b_2 \tau_t^W \left(1 - \exp\left(-\dfrac{|\tau_t^W|}{\sigma_2^2}\right)\right) \\ \hat{I}_{t,j} = f(I_{t,j}, \tau_t^I) = a_3 I_{t,j} + b_3 \tau_t^I \left(1 - \exp\left(-\dfrac{|\tau_t^I|}{\sigma_3^2}\right)\right) \end{cases} \tag{11}$$

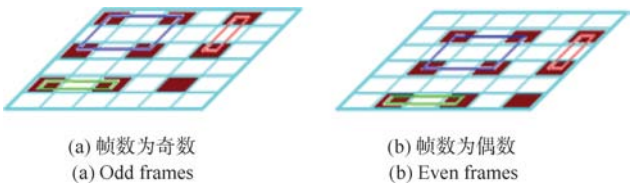
码字中其余参数按下式更新:

$$\begin{cases} f_{t,j} = f_{t,j} + 1 \\ \lambda_{t,j} = \max(\lambda_{t,j}, t - \lambda_{t,j}) \\ q_{t,j} = t \end{cases} \tag{12}$$

在实验过程中, 我们发现参数 a_1, a_2, a_3 取值均为 1, b_1, b_2 取值为 1, b_3 取 0.1, 以及 σ_1, σ_2, σ_3 取值为 525, 425, 625 时, 背景模型适应不同场景的能力较强.

2　分步的融合时空信息的背景模型

为了较好地利用同一帧图像内相邻像素之间的相关性, 我们先将测试图像分成两幅子图, 然后, 用训练得到的时域信息背景模型检测其中一幅子图的状态, 将检测结果作为另一幅子图的先验信息, 同时采用马尔科夫随机场约束该先验信息, 最终检测该幅子图的状态. 为避免某些像素一直提供先验信息, 在采样过程中, 我们交替执行两种采样方式, 即: 奇数行采样与偶数行采样 (如图 4).



(a) 帧数为奇数 (a) Odd frames　　(b) 帧数为偶数 (b) Even frames

图 4　采样结果

Fig. 4　The results of sampling

2.1 邻域系

对测试图像 F_t 进行采样, 当 t 为奇数时, 取原图像中行和列皆为奇数的像素 (如图 4 (a) 中黑色格子) 组成子图 F_1, 当 t 为偶数时, 取原图像中行和列皆为偶数的像素 (如图 4 (b) 中黑色格子) 组成子图 F_1; 其余像素 (如图 4 中的空白格子) 组成子图 F_2.

从图 4 中可以看出, 在不考虑边缘的情况下, 子图 F_1 和子图 F_2 只存在三种邻域关系: 即 F_2 图像像素的上下存在 F_1 图像像素点; F_2 图像像素左右存在 F_1 图像像素点; F_2 图像像素的斜四角存在 F_1 图像像素点. 这三种相邻关系依次如图 4 中的线框覆盖区域所示.

2.2 背景模型构建

子图 F_1 和子图 F_2 的相邻关系可用无向图 $G = (V, E)$ 来表示, 其中, V 为 F_1 中每一像素对应的节点集 V_1 (对应状态变量集为 L_1) 和 F_2 中每一像素对应的节点集 V_2 (对应状态变量集为 L_2) 的并集, 即 $V = V_1 \cup V_2$. E 为连接具有相邻关系节点的边的集合. 我们先利用训练得到的时域信息背景模型检测子图 F_1 中每一像素的状态, 将检测结果作为 F_2 的先验信息, 同时采用马尔科夫随机场约束该先验信息, 以检测子图 F_2 中每一像素的状态.

根据贝叶斯准则 (Bayesian rule), 子图 F_2 的后验概率为

$$p(L_2|F_2) \propto p(L_2 = X)p(F_2|L_2) \tag{13}$$

其中, $p(L_2 = X)$ 为先验概率, 这里 X 为 L_2 的一具体组合, $p(F_2|L_2)$ 为似然函数.

根据马尔科夫随机场理论[13−15], 在本文定义的无向图 $G = (V, E)$ 中, 以节点 $v_i \in V_2$ 为中心的邻域系的先验概率如下:

$$p(l_i^2) \propto \exp\left(-\sum_{v_j \in Ne(v_i)} w_1(1 - \delta(l_i^2 - l_j^1))\right) \tag{14}$$

其中, $l_i^2 \in L_2$ 为该节点对应的状态变量, $l_j^1 \in L_1$ 为其邻域节点的状态变量, w_1 为衡量邻域节点之间连续性的权重因子. 则 F_2 的先验概率为

$$p(L_2 = X) = \prod_{v_i \in V_2} p(l_i^2) \tag{15}$$

记 $\boldsymbol{f}_t^i$ 为第 i 个节点处的观测变量, 则该节点的似然函数为

$$p(\boldsymbol{f}_t^i|l_i^2) \propto \exp(-w_2(1 - \delta(BGS(\boldsymbol{f}_t^i) - l_i^2))) \tag{16}$$

其中, w_2 为权重因子, 它决定着似然函数对后验概率的影响程度, $BGS(\boldsymbol{f}_t^i)$ 表示根据式 (5) 和式 (7) 判断节点 i 是背景还是前景的函数, 若是背景, 其值为 0, 若是前景, 其值为 1. 则 F_2 在 L_2 下的似然函数为

$$p(F_2|L_2) = \prod_{v_i \in V_2} p(\boldsymbol{f}_t^i|l_i^2) \tag{17}$$

根据式 (13), 结合等式 (15) 和 (17), 得到后验概率为

$$p(L_2|F_2) \propto p(L_2 = X)p(F_2|L_2) = \prod_{v_i \in V_2} p(l_i^2) \prod_{v_i \in V_2} p(\boldsymbol{f}_t^i|l_i^2) \tag{18}$$

即判断子图 F_2 中每一像素点的状态等价于求 L_2 的最优组合 $\hat{X}$, 使得后验概率 $p(L_2 \mid F_2)$ 最大.

2.3 背景模型推理

为计算方便, 对式 (18) 两边取负对数, 结合式 (14) 和 (16), 可得后验概率的能量函数为

$$E(L_2|F_2) = \sum_{v_i \in V_2} \sum_{v_j \in Ne(v_i)} w_1(1 - \delta(l_i^2 - l_j^1)) + \sum_{v_i \in V_2} w_2(1 - \delta(BGS(\boldsymbol{f}_t^i) - l_i^2)) \tag{19}$$

那么求解后验概率 $p(L_2|F_2)$ 的最大值即等价于寻找 L_2 的最优组合 $\hat{X}$, 使得能量函数 $E(L_2|F_2)$ 最小. 则对于节点 $v_i \in V_2$ 的局部能量函数为

$$E_{v_i}(L_2|F_2) = \sum_{v_j \in Ne(v_i)} w_1(1 - \delta(l_i^2 - l_j^1)) + w_2(1 - \delta(BGS(\boldsymbol{f}_t^i) - l_i^2)) \tag{20}$$

能量最小化的经典方法有 ICM[16] (Iterated conditional modes)、LBP[17] (Max-product loopy belief propagation)、GC[18−19] (Graph-cut algorithms) 等. 本文构建的马尔科夫随机场 (MRF) 中, 状态变量为二值变量, 且 MRF 的邻域中只有 2 个或 4 个变量, 总体计算复杂度较低. 综合考虑求解最大后验概率时的时效性和简洁性, 本文采用 GC 算法求解.

3 实验结果和分析

为了验证算法性能, 在多个标准测试视频上测试了本文算法 (Our model), 并与 MOG[5] (Mixture of Gaussians)、SCBM[7] (Standard codebook model) 和 ICBM[11] (Improved codebook model) 进行比较.

实验在 Intel(R) C 3.2 GHz 4.0 GB 计算机上在 VC++ 和 Open CV2.1 环境下编程实现. 在对比实

验中, MOG、SCBM 和 ICBM 中的参数采用原文献中的参数. 在实现过程中, 没有对视频进行预处理, 也没有对实验结果进行后处理, 如形态学操作.

3.1 实验结果

实验 1. 光照突变和更新策略

本测试视频来自 Pets2001DataSet3, 共 5 336 帧, 整个视频中多处存在光照突然变化. 图 3 给出了背景中墙面上 4 个样本点 (a, b, c, d) 在前 3 010 帧图像中随场景光照变化时的亮度以及色度变化情况. 从图 3 (b) ∼ (c) 可以看出, 大约在第 600 帧到 720 帧之间, 样本点所在区域存在光照突然变化, 从而使得亮度发生了剧烈变化, 同时黄蓝通道色度也发生

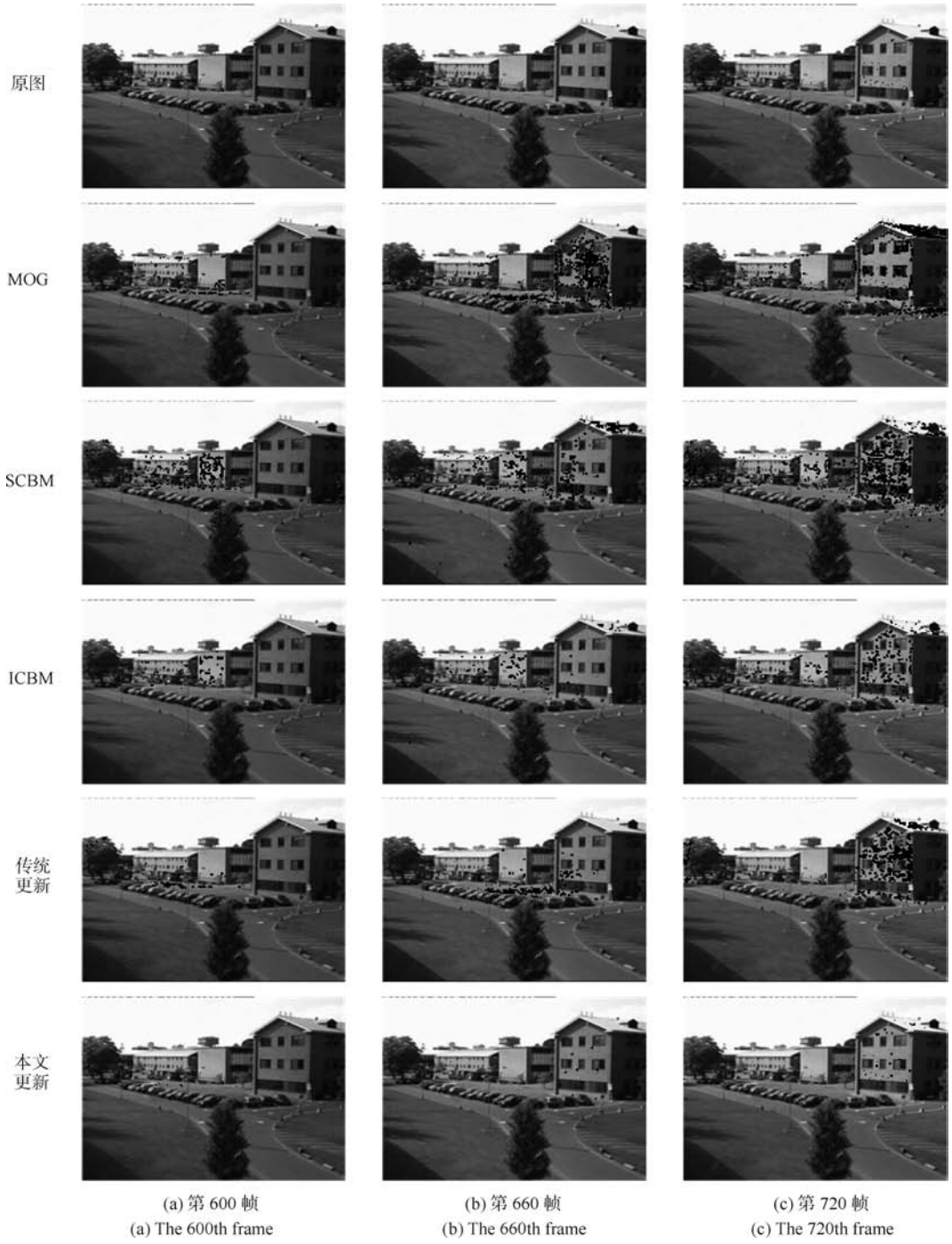


图 5　验证光照突变和更新策略 (其中深色为检测前景)

Fig. 5　Test illumination changes and updating strategy (The black is detection foreground.)

了变化, 红绿通道色度出现了微小的变化. 为了验证本文算法能较好应对场景光照突然变化, 我们给出了这段范围内的实验结果, 如图 5 所示.

实验 1 (a). 光照突变时采用传统更新策略

图 5 中第一行为原始视频; 第 2～4 行分别为 MOG、SCBM 和 ICBM 的检测结果, 可以看出随光照变化, 相对于 ICBM、MOG 和 SCBM 虚警数 (False positive, FP)[20] 较高; 第 5 行为本文算法检测的结果, 但是背景模型的更新方法采用的是 SCBM 的方法, 尽管本文算法采用了具有光照不变性的颜色空间来表征背景, 以及较充分地利用了同一帧图像内相邻像素的相关性, 但是如图 3 (d) 所示, 在这段范围内黄蓝通道色度出现了变化, 而传统的更新方法难以捕捉这种变化, 因此, 本文算法的虚警数 FP 相对于 ICBM 也较高, 但还是比 MOG 以及 SCBM 低.

实验 1 (b). 光照突变时采用本文更新策略

图 5 中第 6 行也为本文算法检测的结果, 但这里采用的是本文提出的背景更新策略来更新背景. 从实验结果可以看出, 采用本文提出的背景更新策略动态更新背景后, 相对于第 5 行的检测结果, 虚警数 FP 很大程度上得到减少. 可以看出本文提出的背景更新策略增强了本文背景模型应对场景出现光照变化时的适应性.

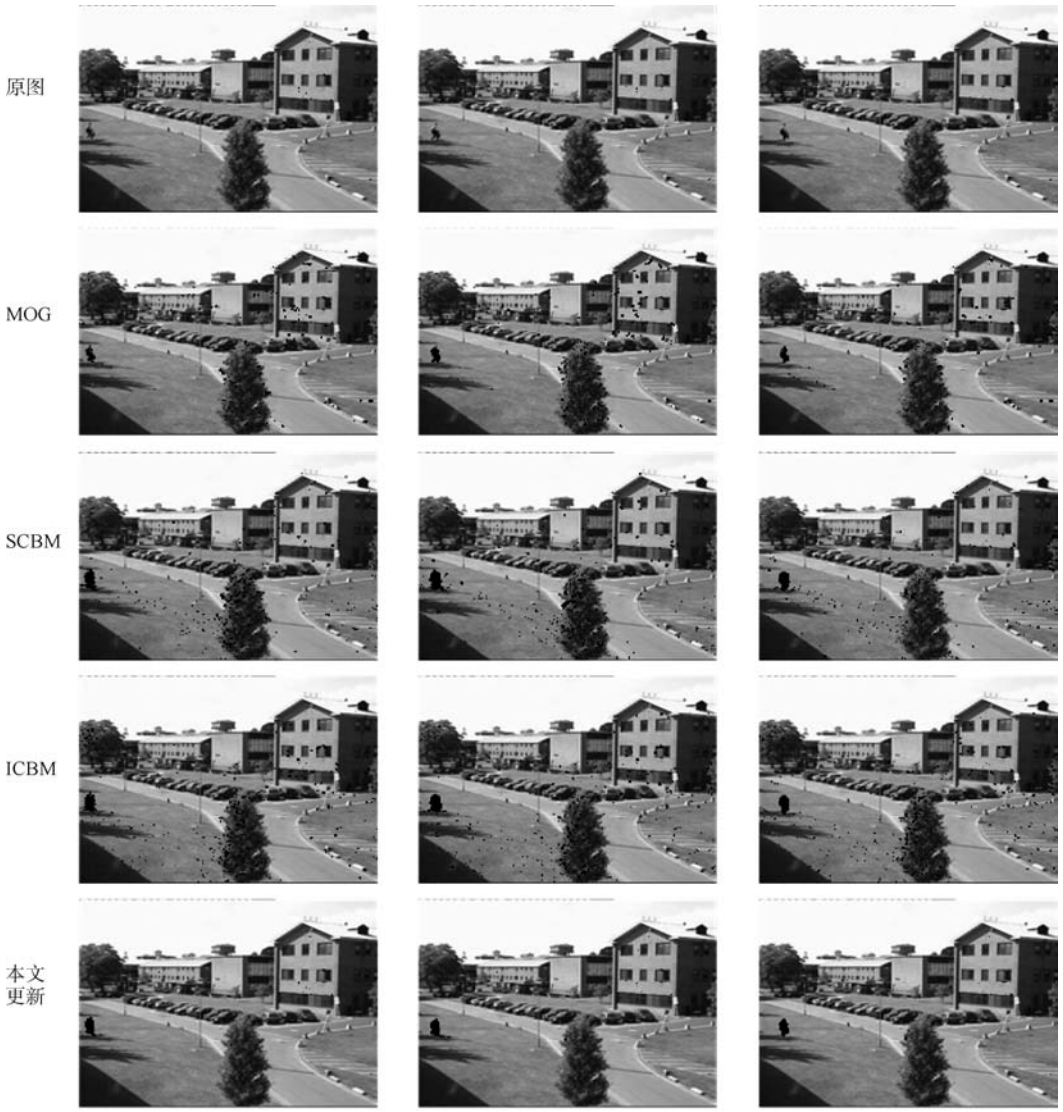


图 6 检测运动物体 (其中深色为检测前景)

Fig. 6 Detection moving object (The black is detection foreground.)

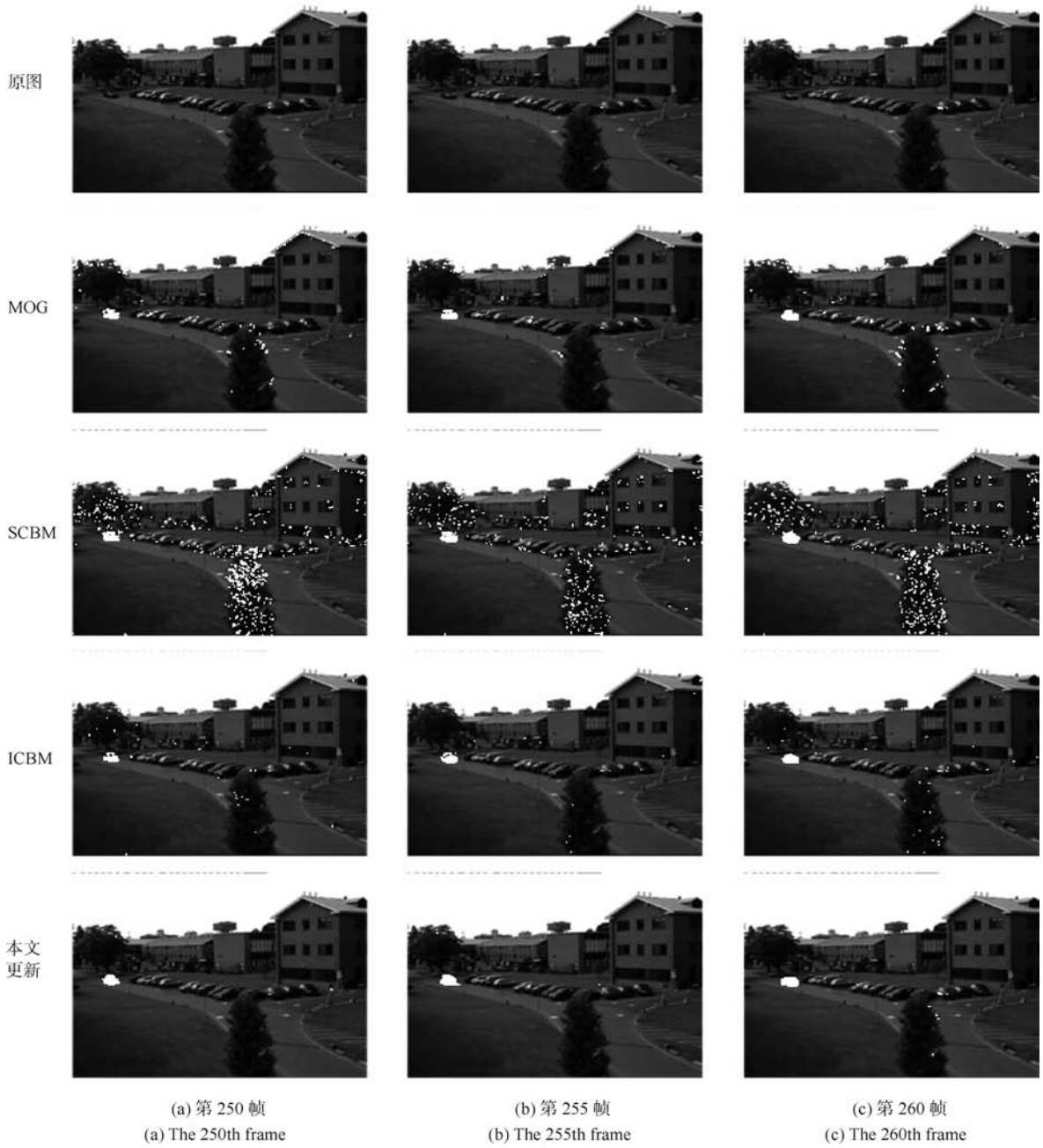


图 7　关于噪声的试验 (其中白色为检测前景)

Fig. 7　Experiment about noise (The white is detection foreground.)

实验 2. 运动物体检测

测试视频还是 Pets2001DataSet3. 实验结果如图 6 所示, 可以看出在漏检数 (False negative, FN)[20] 大致相同的情况下, SCBM 和 ICBM 的虚警数 FP 较高, 其次是 MOG, 而本文算法的虚警数 FP 是最少的.

实验 3. 关于噪声的实验

本测试视频来自 Pets2001DataSet2, 该视频序列共有 2 823 帧, 在第 250 帧到 260 帧间, 背景比较稳定, 几乎不存在光照变化和树枝的不规则运动, 但存在人为添加的高斯噪声. 实验结果如图 7 所示, 可以看出 4 种背景模型中 SCBM 出现的虚警数 FP 最多, 原因在于计算颜色失真度时, 颜色失真度会因噪声存在出现波动, 当场景中照度较低时, 这种波动会更为明显, 因此, SCBM 在背景较暗的地方出现了较高的误检测率, MOG 在应对高斯噪声时效果明显比前者好, ICBM 由于利用了相邻帧以及当前帧的空域的相关性, 情况有进一步的改善, 但是在虚

警数 FP 几乎一致的情况下, 本文算法比 ICBM 的漏检数 FN 更低, 可见本文的码字聚类准则对噪声有较好的鲁棒性.

实验 4. 树枝不规则运动

测试视频 Wavetrees 一共 286 帧, 该视频背景中的树枝存在剧烈的不规则运动, 另一方面由于整个背景的大半部分被树枝占据, 更增加了其检测运动前景的难度. 实验结果如图 8, 可以看出 4 种算法中 MOG 的虚警数 FP 和漏检数 FN 都较高, 此外, SCBM 的虚警数 FP 也较高, ICBM 的虚警数 FP 虽稍微少于本文算法, 但是本文算法的漏检数 FN 比 ICBM 低.

实验 5. 水面纹理

测试视频 Water 共 291 帧, 背景 – 水面波涛起伏, 另有雨点打击水面, 使得背景的不规则运动更加复杂. 实验结果如图 9. 其中 SCBM 的虚警数 FP 最高, 这些误检测有两个来源: 水面波纹和漂浮于水面的瓶子阴影处的误检测. 相对于 SCBM, ICBM 对水面波纹的虚警数 FP 有所降低, 但是在瓶子的阴影处, 误检测仍然较多, MOG 整体效果较之前两种算法更好. 而本文算法在这两处的虚警数 FP 都有所降低, 因此, 在漏检数 FN 大致相同的情况下, 本文的虚警数 FP 最少.

3.2 性能评估

本文先采用虚警数 FP 和漏检数 FN 来定量评估算法性能, 其中, 虚警数是指背景被误检测为前景的像素个数, 漏检数是指前景被误检测为背景的像

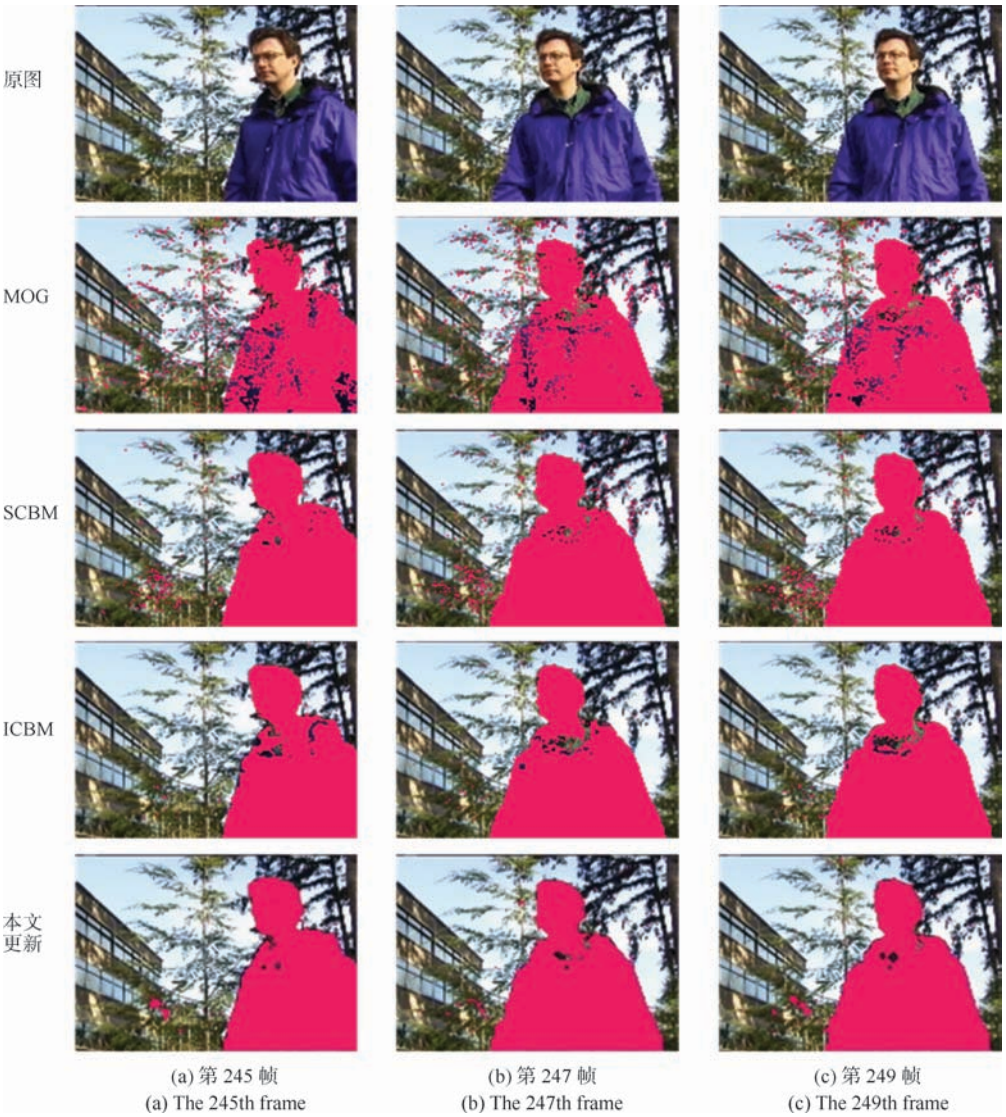


图 8 关于树枝不规则运动的实验 (其中灰色为检测前景)

Fig. 8 Experiment about irregular movement of tree (The gray is detection foreground.)

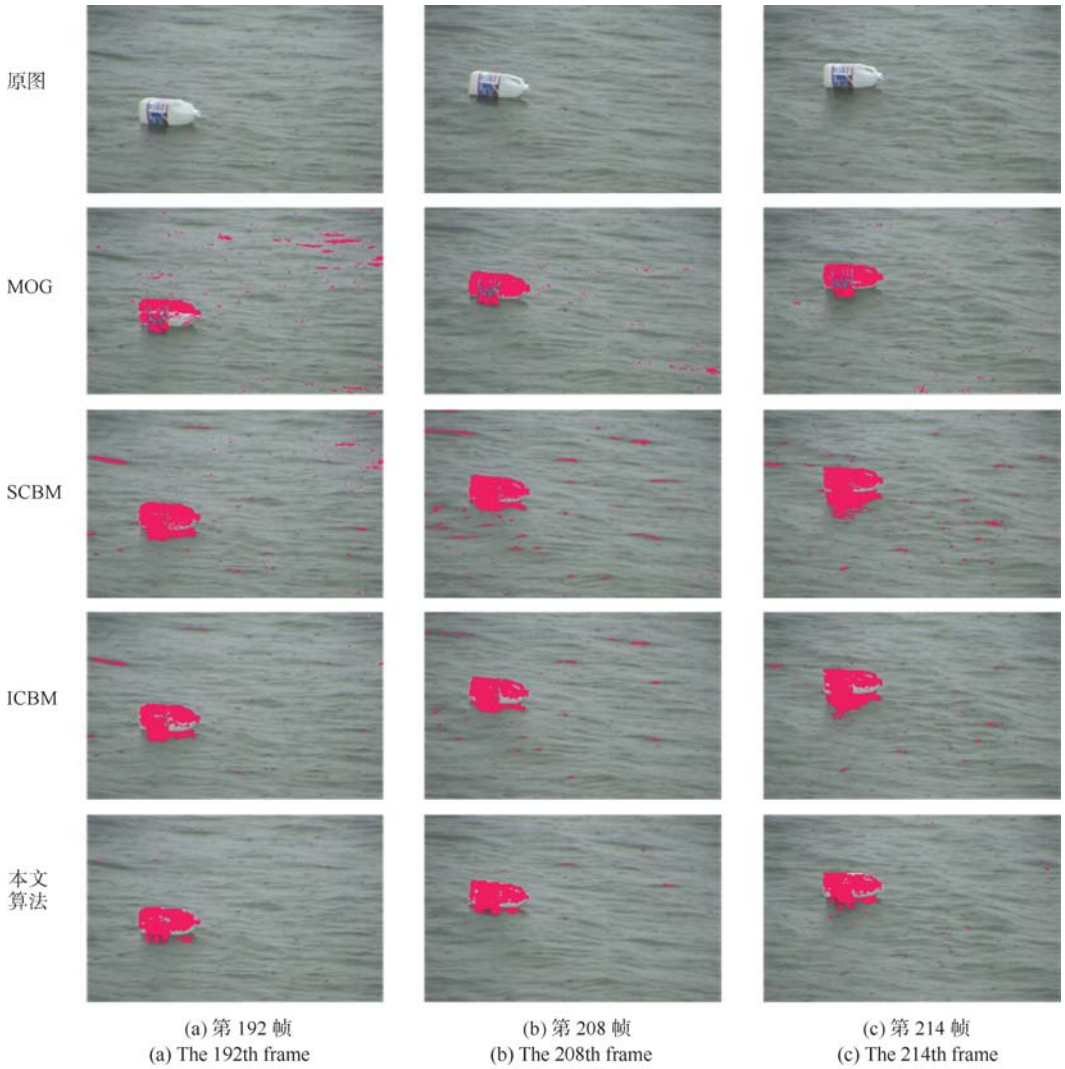


图 9　关于水面纹理的实验 (其中深色为检测前景)

Fig. 9　Experiment about texture of water surface (The black is detection foreground.)

素个数. 以上几组实验的虚警数 FP 和漏检数 FN 以及它们综合贡献的结果如图 10 (实验 1 中没有运动前景, 不存在漏检数 FN), 从图 10 (c) 可以看出, 相对于 MOG、SCBM 和 ICBM, 本文算法总体上出现的误检测最少. 我们也采用了文献 [21] 中的相似度对算法做综合评定. 相似度为 $S(D_t, G_t) = |D_t \cap G_t| / |D_t \cup G_t|$, 其中 D_t 为检测的运动前景区域, G_t 为手工标注的运动前景区域. 以上几组实验的相似度如表 1 所示. 可以看出本文算法比前三种对比算法, 相似度有了较大提高.

3.3　耗时分析

在相同的测试平台下, 我们将 MOG、SCBM、ICBM 以及本文算法的处理速度进行了比较. 其中, 除 Wavetrees 的尺寸为 160×120 外, 其余视频的大小均为 320×240, 实验结果如表 2 所示. 从表 2 中可以看出 ICBM 运行速度最慢, MOG 运行速度最快, 其次是 SCBM, 而本文算法比 SCBM 稍慢, 却基本上可以达到实时, 但是本文的检测精度总体来说较高.

表 1　相似度值

Table 1　The similarity degree

	MOG	SCBM	ICBM	Our method
实验 2	0.2080	0.1832	0.1604	0.7177
实验 3	0.3214	0.1640	0.5634	0.7671
实验 4	0.8102	0.9137	0.9402	0.9432
实验 5	0.5508	0.5042	0.5397	0.6934
平均值	0.4726	0.4413	0.5509	0.7804

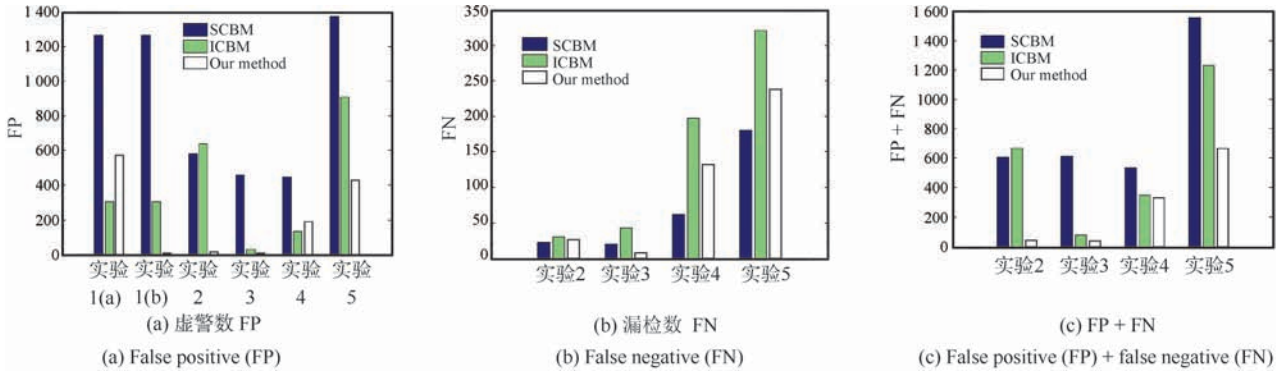


图 10 虚警数 FP 和漏检数 FN

Fig. 10 False positive and false negative

表 2 算法处理速度 (fps)

Table 2 Processing times of algorithm (fps)

	实验 2	实验 3	实验 4	实验 5
MOG	21.09	21.76	62.33	26.52
SCBM	20.27	18.95	54.93	26.31
ICBM	6.90	6.11	17.17	8.49
Our method	19.30	17.50	53.82	25.24

4 结论

针对自然场景中存在光照突然变化和不规则运动的问题, 提出了一种分步的融合时域信息和空域信息的背景建模方法. 由于在光照不变性的颜色空间下构造码字, 并且在码字聚类时考虑噪声的分布和光照的变化, 以及采用了自适应的背景更新方法, 因此, 本文算法对噪声和光照突然变化具有较好的适应性; 采用分步的策略、利用马尔科夫随机场将背景中的时域信息和空域信息有机融合, 能较好地克服场景中不规则运动的影响. 采用多个标准视频进行测试, 并与 MOG、SCBM 和 ICBM 进行比较, 结果表明本文算法具有较强的背景描述能力.

References

1 Qi Yu-Juan, Wang Yan-Jiang, Li Yong-Ping. Memory-based Gaussian mixture background modeling. *Acta Automatica Sinica*, 2010, **36**(11): 1520−1526
(齐玉娟, 王延江, 李永平. 基于记忆的混合高斯背景建模. 自动化学报, 2010, **36**(11): 1520−1526)

2 Wang Yong-Zhong, Liang Yan, Pan Quan, Cheng Yong-Mei, Zhao Chun-Hui. Spatiotemporal background modeling based on adaptive mixture of Gaussians. *Acta Automatica Sinica*, 2009, **35**(4): 371−378
(王永忠, 梁彦, 潘泉, 程咏梅, 赵春晖. 基于自适应混合高斯模型的时空背景建模. 自动化学报, 2009, **35**(4): 371−378)

3 Wren C R, Azarbayejani A, Darrell T, Pentland A P. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780−785

4 Horprasert T, Harwood D, Davis L S. A statistical approach for real-time robust background subtraction and shadow detection. In: Proceedings of the 12th International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 1−19

5 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the 14th International Conference on Computer Vision and Pattern Recognition. Fort Collins, USA: IEEE, 1999. 246−252

6 Chen G, Yu Z Z, Wen Q, Yu Y Q. Improved Gaussian mixture model for moving object detection. *Artificial Intelligence and Computational Intelligence*, 2011, **7002**: 179−186

7 Kim K, Chalidabhongse T H, Harwood D, Davis L. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 2005, **11**(3): 172−185

8 Zhao Xu-Dong, Liu Peng, Tang Jiang-Long, Liu Jia-Feng. Background modeling adaptive to outdoor illumination variation and foreground detection approach. *Acta Automatica Sinica*, 2011, **37**(8): 915−922
(赵旭东, 刘鹏, 唐降龙, 刘家锋. 一种适应户外光照变化的背景建模及目标检测方法. 自动化学报, 2011, **37**(8): 915−922)

9 Yin Z Z, Collins R. Belief propagation in a 3D spatio-temporal MRF for moving object detection. In: Proceedings of the 23rd Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE, 2007. 1−8

10 Migdal J, Grimson W E L. Background subtraction using Markov thresholds. In: Proceedings of the 7th Workshop on Application of Computer Vision. Breckenridge, USA: IEEE, 2005. 58−65

11 Wu M J, Peng X R. Spatio-temporal context for codebook-based dynamic background subtraction. *AEU-International Journal of Electronics and Communication*, 2010, **64**(8): 739−747

12 Burghouts G J, Geusebroek J M. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 2009, **113**(1): 48−62

13 Li S Z. Markov random field modeling in computer vision. In: Proceedings of the 3rd European Conference on Computer Vision. Secaucus, NJ: Springer-Verlag, 1995. 361−370

14 Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 1974, **36**(2): 192−236

15 Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(6): 1068−1080

16 Besag B J. On the statistical analysis of dirty picture. *Journal of the Royal Statistical Society, Series B (Methodological)*, 1986, **48**(3): 259−302

17 Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 2006, **70**(1): 41−54

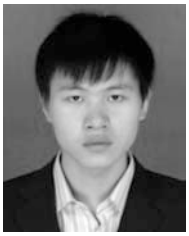
18 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(11): 1222−1239

19 Tao Lin-Mi, Wang Qi-Fan, Di Hui-Jun. Markov random field in visual information processing. *Journal of Image and Graphics*, 2009, **14**(9): 1705−1711
(陶霖密, 王奇凡, 邸慧军. 视觉信息处理中的马尔可夫随机场. 中国图象图形学报, 2009, **14**(9): 1705−1711)

20 Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4): 657−662

21 Chen Y T, Chen C S, Huang C R, Huang Y P. Efficient hierarchical method for background subtraction. *Pattern Recognition*, 2007, **40**(10): 2706−2715

储 珺 南昌航空大学计算机视觉研究所教授. 主要研究方向为图像处理与计算机视觉. 本文通信作者.
E-mail: chujun99602@163.com
(**CHU Jun** Professor at Institute of Computer Vision, Nanchang Hangkong University. Her research interest covers image processing and computer vision. Corresponding author of this paper.)



杨 樊 南昌航空大学计算机视觉研究所硕士研究生. 主要研究方向为目标检测与智能视频监控.
E-mail: 693339173@qq.com
(**YANG Fan** Master student at Institute of Computer Vision, Nanchang Hangkong University. His research interest covers visual object detecting and intelligent video surveillance.)

张桂梅 南昌航空大学计算机视觉研究所教授. 主要研究方向为图像处理, 计算机视觉与模式识别.
E-mail: guimei.zh@163.com
(**ZHANG Gui-Mei** Professor at Institute of Computer Vision, Nanchang Hangkong University. Her research interest covers image processing, computer vision, and pattern recognition.)

汪凌峰 中国科学院自动化所助理研究员. 主要研究方向为计算机视觉和图像处理.
E-mail: lfwang@nlpr.ia.ac.cn
(**WANG Ling-Feng** Assistant researcher at the Institute of Automation, Chinese Academy of Sciences. His research interest covers computer vision and image processing.)