

面向运动目标检测的粒子滤波视觉注意力模型

刘 龙,樊波阳,刘金星,杨乐超

(西安理工大学自动化学院,陕西西安 710048)

摘 要: 视觉注意力是机器视觉领域的研究热点,对目标检测、跟踪等技术发展具有积极意义,本文面向运动目标检测问题,构建了一种基于粒子滤波的视觉注意力模型. 首先依据贝叶斯估计理论,推导了基于注意力的粒子权重计算方法;然后将运动注意力和目标颜色注意力分别作为自底向上(Bottom-Up)和自顶向下(Top-Down)注意力的输入,通过重要性采样、粒子权重计算、重采样等形成粒子注意力显著图,并确定目标位置;测试结果显示本文方法能够获取比其它方法更好的目标注意力显著图,并具有准确的目标检测效果.

关键词: 注意力模型; 运动目标检测; 粒子滤波; 融合

中图分类号: TN911. 7 **文献标识码:** A **文章编号:** 0372-2112 (2016)09-2235-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2016.09.031

Particle Filtering Based Visual Attention Model for Moving Target Detection

LIU Long, FAN Bo-yang, LIU Jin-xing, YANG Le-chao

(Xi'an University of Technology, Xi'an, Shaanxi 710048, China)

Abstract: Visual attention is one of the research hotspots in the field of machine vision, which is positive significance for development of target detection and target tracking. This paper presents a particle filter based visual attention model that is applied to detect moving target. Firstly, according to Bayes estimation theory, the method of particle weight calculation is established by visual bidirectional (Top-Down/Bottom-Up) fusion attention. Then motion attention and target color attention are adopted as input of the attention model, and moving target saliency is calculated through the importance sampling, particle weight calculation, resampling and particle saliency map processing. Lastly, the target position is determined by distribution of particle. Different video complex scene test results show that this method is more effective and accurate than the traditional method for detection of moving target.

Key words: motion attention model; moving target detection; particle filtering; fusion

1 引言

运动目标检测是机器视觉领域的重要问题之一,是目标跟踪、识别的前提,但在复杂运动场景中,现有运动目标检测方法仍存在一定局限和不足. 传统的运动目标检测方法有背景差分法、帧间差分法和全局运动补偿法,背景差分法和帧间差分法只局限于局部运动场景,全局运动补偿法适用范围广,但准确性受到目标大小和运动强度影响,当目标体较大或运动较强烈时,全局运动估算准确性下降,不能进行全局有效补偿,导致运动目标检测存在较大误差. 近些年,视觉感知研究逐渐融入了人类生理和心理的研究成果,视觉注意力研究是其中典型问题之一,其主要研究如何通过计算

机模拟人类视觉系统对外界的初级反应,对目标检测、跟踪等问题具有积极意义.

人类视觉注意力是由自底向上(Bottom-Up, B-U)和自顶向下(Top-Down, T-D)两种注意力共同作用产生的. 2002年,Itti和Koch^[1]建立了以亮度、颜色等特征为主的B-U注意力模型,随后的一些研究衍生出多种视觉注意力计算方法及应用. 目前,视觉注意力模型大致可分为自底向上型^[1-5]和双向结合型^[6-10]两种类型. 自底向上型描述B-U注意力计算方法;双向结合型描述B-U和T-D注意力共同作用的计算方法.

视觉注意力模型构建有两种思路,一种是以模拟人类视觉机理为主要目的,另一种是针对一些具体问题构建视觉注意力模型,这类研究以解决实际问题为

主要目标. 近些年, 一些学者从注意力的角度对目标检测问题进行了一定的研究, 文献[9, 11 ~ 12]采用双向注意力方法进行静态目标检测. Fang Y M等^[9]提取目标的方向特征作为 T-D 注意力, 并与 B-U 注意力进行比例加权融合, 最终确定目标位置. Alcides X. B 等对特定静态目标特征进行自组织神经网络学习, 生成权重矩阵, 并将其作为自顶向下的影响因素调节 B-U 注意力计算过程, 形成目标注意力显著图. Yu Y L^[11, 12]等建立目标特征长期记忆(LTM)单元, 通过与低级特征的对比计算位置概率分布偏置, 并进行双向加权融合, 确定目标位置. 另外, 文献[13]建立了运动注意力模型进行运动目标检测, 主要思路是根据运动反差定义运动注意力模型, 使运动显著性区域逼近目标区域. Ma Y F^[13]根据从 MPEG 码流中解压所得到的运动矢量场的运动矢量能量、空间相关性和时间相关性综合定义了运动注意力模型, 通过该模型可以得到运动显著性区域. Han J W^[14]将注意力分为静态注意力和动态注意力两种, 静态注意力主要由图像的亮度、颜色等信息引起, 动态注意力是在全局运动补偿的基础上计算区域的变化像素的比例进行定义的, 注意力模型最终由静态注意力和动态注意力融合得到, 主要应用于运动目标检测.

文献[11 ~ 13]涉及了目标检测问题, 但主要是解决静态目标的注意力问题, 采用静态目标特征如轮廓、颜色等, T-D 和 B-U 注意力采用了加权融合, 较难反应两者对最终目标注意力的作用, 文献[13 ~ 14]根据运动信息构建运动注意力模型, 属于 B-U 注意力计算方法, 具有一定的局限性. 综上所述, 相关研究还未有针对运动目标检测的注意力模型, 而现有注意力模型缺乏将运动注意力以及目标特征等综合计算的形式. 针对上述问题, 本文着重研究面向运动目标检测的视觉注意力方法, 提出借鉴粒子滤波机制建立双向融合注意力模型, 采用运动注意力、目标颜色注意力分别作为 B-U 和 T-D 双向注意力输入, 通过滤波后的粒子分布状态计算注意力显著度, 并最终确定目标位置.

2 粒子滤波双向融合注意力模型

根据粒子滤波理论, 设粒子状态为 $\{X^{(i)}\}_{i=1}^N$, 观测值为 Z^k , 则 k 时刻的后验概率密度函数可近似为:

$$p(\mathbf{X}^k | \mathbf{Z}^k) \approx \sum_{i=1}^N \lambda^{(i)} \cdot \delta(\mathbf{X}^k - \mathbf{X}_{(i)}^k) \quad (1)$$

其中, $\lambda^{(i)} \propto \frac{\pi(\hat{x}^{(i)})}{q(\hat{x}^{(i)})}$, $i = 1, 2, \dots, N$. 假设双向融合注意力显著图状态为 $x_{0:t}$, B-U 和 T-D 注意力观测值分别为 $z_{1:t}^{\text{T-D}}$ 和 $z_{1:t}^{\text{B-U}}$, 则后验概率 $p(\mathbf{X}^T | \mathbf{Z}^T)$ 可表示为 $p(x_{0:t} | z_{1:t}^{\text{B-U}}, z_{1:t}^{\text{T-D}})$, 双向注意力的后验概率值求解推导如下:

$$p(x_{0:t} | z_{1:t}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) = \frac{p(z_{1:t}^{\text{B-U}} | x_{0:t}, z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) \cdot p(x_{0:t} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})}{p(z_{1:t}^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})} \quad (2)$$

$$= \frac{p(z_t^{\text{B-U}} | x_t, z_{1:t}^{\text{T-D}}) \cdot p(x_{0:t} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})}{p(z_t^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})} \quad (3)$$

$$= \frac{p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(z_t^{\text{B-U}} | x_t)}{p(z_t^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) \cdot p(z_t^{\text{T-D}} | x_t)} \cdot \frac{p(z_t^{\text{T-D}} | x_{0:t}, z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \cdot p(x_{0:t} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}})}{p(z_t^{\text{T-D}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})} \quad (4)$$

$$= \frac{p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(z_t^{\text{B-U}} | x_t)}{p(z_t^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) \cdot p(z_t^{\text{T-D}} | x_t)} \cdot p(x_{0:t-1} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \cdot \frac{p(z_t^{\text{T-D}} | x_{0:t}, z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \cdot p(x_t | x_{0:t-1}, z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}})}{p(z_t^{\text{T-D}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})} \quad (5)$$

$$= \frac{p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(z_t^{\text{B-U}} | x_t)}{p(z_t^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) p(z_t^{\text{T-D}} | x_t)} \cdot \frac{p(x_t, z_t^{\text{T-D}} | x_{0:t-1}, z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}})}{p(z_t^{\text{T-D}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}})} \cdot p(x_{0:t-1} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \quad (6)$$

$$= \frac{1}{k_t} \cdot p(z_t^{\text{B-U}} | x_t) \cdot p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(x_t, z_t^{\text{T-D}} | x_{0:t-1}) \cdot \frac{1}{p(z_t^{\text{T-D}} | x_t)} \cdot p(x_{0:t-1} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \quad (7)$$

$$= \frac{1}{k_t} \cdot p(z_t^{\text{B-U}} | x_t) \cdot p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(x_t | x_{0:t-1}) \cdot p(z_t^{\text{T-D}} | x_t, x_{0:t-1}) \cdot \frac{1}{p(z_t^{\text{T-D}} | x_t)} \cdot p(x_{0:t-1} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \quad (8)$$

$$= \frac{1}{k_t} \cdot p(z_t^{\text{B-U}} | x_t) \cdot p(z_t^{\text{T-D}} | x_t, z_t^{\text{B-U}}) \cdot p(x_t | x_{0:t-1}) \cdot p(x_{0:t-1} | z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) \quad (9)$$

上述推导遵循以下假设条件: (i) 时间动态过程符合马尔科夫过程; (ii) 不同时间的观测值相互独立, 观测值只与当前状态相关. 根据假设(i)和假设(ii), 对推导步骤作了如下简化:

$$(1) \text{ 从式(2)到式(3)推导步骤中做了简化 } p(z_t^{\text{B-U}} | x_{0:t}, z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) = p(z_t^{\text{B-U}} | x_t, z_{1:t}^{\text{T-D}});$$

$$(2) \text{ 从式(6)到式(7)推导步骤里简化 } p(x_t, z_t^{\text{T-D}} | x_{0:t-1}, z_{1:t-1}^{\text{B-U}}, z_{1:t-1}^{\text{T-D}}) = p(x_t, z_t^{\text{T-D}} | x_{0:t-1});$$

$$(3) \text{ 式(6)的分母, 密度 } p(z_t^{\text{B-U}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) \text{ 和 } p(z_t^{\text{T-D}} | z_{1:t-1}^{\text{B-U}}, z_{1:t}^{\text{T-D}}) \text{ 与 } x_t \text{ 无关, 可以认为常量 } k_t;$$

$$(4) \text{ 式(8)中简化 } p(z_t^{\text{T-D}} | x_t, x_{0:t-1}) = p(z_t^{\text{T-D}} | x_t).$$

根据重要性采样定理, 粒子权重 $\lambda^{(i)}$ 正比 $\frac{\pi(\hat{x}_{0:t}^{(i)} | z_{0:t})}{q(\hat{x}_{0:t}^{(i)} | z_{0:t})}$, 根据式(9)可得:

$$\lambda_t^{(i)} \approx \lambda_{t-1}^{(i)} \cdot p(z_t^{\text{B-U}} | \hat{x}_t^{(i)}) \cdot p(z_t^{\text{T-D}} | \hat{x}_t^{(i)}, z_t^{\text{B-U}}) \quad (10)$$

3 基于视觉注意力的运动目标检测

在复杂运动场景中,目标运动与全局运动形成运动反差,这种反差形成 B-U 运动注意力,其显著度记作 $SM^{\text{B-U}}$;目标本身还具备颜色等先验知识特征,可形成 T-D 注意力,其显著度记作 $SM^{\text{T-D}}$. 根据第 2 节中的粒子滤波双向融合注意力模型,将 $SM^{\text{B-U}}$ 和 $SM^{\text{T-D}}$ 作为模型输入值,进行运动目标的检测,其过程为:首先估算当前 t 时刻的运动矢量场 MVF_t ,并进行预处理,再计算 $SM_t^{\text{B-U}}$;然后通过 $SM_t^{\text{B-U}}$ 控制粒子重要性采样;根据目标特征计算 $SM_t^{\text{T-D}}$,并更新粒子权值;重采样后形成新的粒子注意力显著图 SM'_t ;最后,计算粒子分布的重心位置进行目标定位,算法整体框架如图 1 所示.

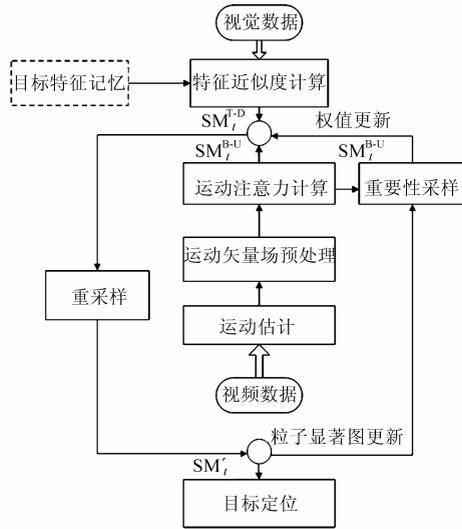


图1 运动目标检测算法框架

3.1 粒子权值计算

本文对运动矢量场进行叠加和滤波两个预处理,运动注意力 $SM^{\text{B-U}}$ 由时间和空间两方面注意力构成,设 S_1 和 S_2 分别表示时间和空间注意力,其计算定义如下:

$$S_1 = |\Delta V| = |V_{t,i,j} - V_{t-1,i,j}| \quad (11)$$

$$S_2 = |V_{t,i,j} - u_{t,i,j}| \quad (12)$$

式中, $V_{t,i,j}$ 和 $V_{t-1,i,j}$ 分别代表在 t 和 $t-1$ 时刻运动矢量场中坐标位置为 (i,j) 的运动矢量, $u_{t,i,j} = \frac{\sum_{\{i,j | V_{t,i,j} \in A_{t,i,j}\}} V_{t,i,j}}{n}$

代表 $V_{t,i,j}$ 邻域 A 范围内的运动矢量均值. 运动注意力由时间和空间注意力线性融合得到,计算如下:

$$SM_t^{\text{B-U}} = \alpha \cdot S_1 + \beta \cdot S_2 \quad (13)$$

式中, α, β 为正值的系数. 图 2 显示了 B-U 注意力显著图.

$SM^{\text{T-D}}$ 的大小采用目标特征与图像特征的相似性程

度衡量,设定颜色直方图作为目标特征量化表示方式,记作为 $\bar{q} = \{\hat{q}^{(u)}\}_{u=1,\dots,m}$, m 为分量个数,则粒子目标区域的颜色分布定义为:

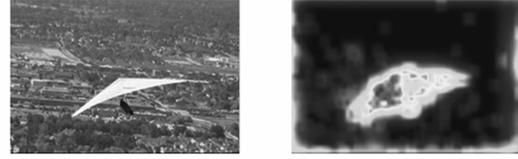


图2 B-U注意力显著图

$$\hat{p}^{(u)}(X_t) = C \cdot \sum_{i=1}^M K\left(\left\|\frac{x - x_i}{h}\right\|\right) \cdot \delta(b(x_i) - u) \quad (14)$$

其中 $\delta(\cdot)$ 为 Delta 函数, $C = 1 / \sum_{i=1}^M K\left(\left\|\frac{x - x_i}{h}\right\|\right)$, 为归一化因子,使得 $\sum_{n=1}^m p_n(y) = 1$, $K(\cdot)$ 为 Epanechnikov 核函数,定义如下:

$$K_E(\|r\|) = \begin{cases} 1 - \|r\|^2, & \|r\| < 1 \\ 0, & \|r\| \geq 1 \end{cases} \quad (15)$$

$SM^{\text{T-D}}$ 显著性计算定义如下:

$$SM^{\text{T-D}} = \sqrt{1 - \rho[\hat{p}(X_t), \bar{q}]} \quad (16)$$

式中, ρ 为巴查理亚 (Bhattacharyya) 系数. 图 3 示显示了 T-D 注意力显著图.

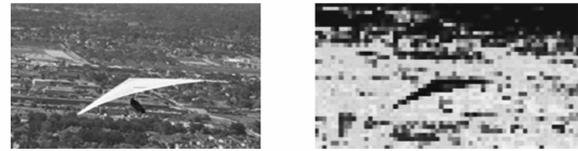


图3 T-D注意力显著图

根据式(11),粒子当前状态权值 $\lambda_t^{(i)}$ 与前力观测条件下的 T-D 注意力后验概率取值,分别定义如下:

$$P(z_t^{\text{B-U}} | \bar{x}_t^{(i)}) = \frac{1}{\sqrt{2\pi} \cdot \delta_{\text{B-U}}} \cdot \exp\left(-\frac{(1 - SM^{\text{B-U}})^2}{2 \cdot \delta_{\text{B-U}}^2}\right) \quad (17)$$

$$P(z_t^{\text{T-D}} | \bar{x}_t^{(i)}, z_t^{\text{B-U}}) = \frac{1}{\sqrt{2\pi} \cdot \delta_{\text{T-D}}} \cdot \exp\left(-\frac{(1 - SM^{\text{T-D}} \cdot e^{SM^{\text{B-U}}})^2}{2 \cdot \delta_{\text{T-D}}^2}\right) \quad (18)$$

3.2 重要性采样与重采样

采用运动显著性特征来调节高斯随机粒子采样的密度,获得随运动显著性变化的随机采样结果,采用高斯随机采样,设 $\bar{z}_1^{(i)} \sim U[0, 1]$, $\bar{z}_2^{(i)} \sim U[0, 1]$, $i = 1, 2, \dots, N$ 独立同分布,令

$$\begin{cases} \bar{x}^{(i)} = \left(\sum_{i=1}^N z_1^{(i)} - N \cdot \mu_x \right) / \sqrt{N \cdot \delta_x^2}, i = 1, 2 \dots N \\ \bar{y}^{(i)} = \left(\sum_{i=1}^N z_2^{(i)} - N \cdot \mu_y \right) / \sqrt{N \cdot \delta_y^2} \end{cases} \quad (19)$$

其中, μ_x, μ_y, δ_x^2 和 δ_y^2 分别是 $\bar{x}^{(i)}, \bar{y}^{(i)}$ 伪随机序列的均值和方差. 通过式(21)可在以 (μ_x, μ_y) 为坐标中心的区域内产生随机的高斯采样结果, 区域内采样粒子密度受运动显著度调控, 假定 $SM_t^{B-U}(x, y)$ 是 t 时刻的显著图在 (x, y) 坐标上的显著值, 定义采样密度函数如下:

$$\Gamma_{(x,y) \in A}(SM^{B-U}) = \frac{1}{\sqrt{2\pi}\delta_A} \cdot \exp\left(-\frac{1}{2} \left(\frac{SM^{B-U}(x_0, y_0) - u_A}{\delta_A} \right)^2\right) \quad (20)$$

式中: 均值 $u_A = \frac{\sum_{(i,j) \in SM^T} SM(i,j)}{N}$, 方差 $\delta_A^2 = \frac{\sum_{(i,j) \in SM^T} [SM(i,j) - u]^2}{N}$, i 和 j 分别表示显著图中的

纵横坐标, 图4显示了粒子采样结果. 粒子权值重新计算后, 加入重采样以淘汰权值较低的粒子, 使粒子向权值高的粒子周围聚集, 本文采用文献[22]中的重采样方法.

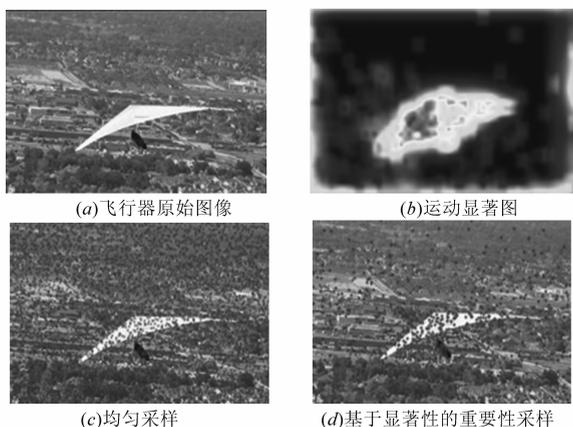


图4 粒子采样

3.3 粒子显著图与目标定位

经过重采样后, 粒子分布的疏密程度反映了注意力强弱, 粒子分布密集区域注意力显著性强, 分布稀疏的区域注意力显著性弱. 根据粒子分布状态, 在二维空间上, 可定义注意力显著大小如下:

$$SM(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma^2} \cdot \varphi\left(\frac{x-x_i}{\sigma}\right) \cdot \varphi\left(\frac{y-y_i}{\sigma}\right) \quad (21)$$

其中, (x, y) 为粒子分布的空间位置, n 为粒子个数, $\sigma = h_n / \sqrt{n}$ 为窗口宽度. 窗函数采用二维高斯窗函数, 那么式(21)变为式(22):

$$SM(x, y) = \frac{1}{2\pi n \sigma^2} \sum_{i=1}^n \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right) \quad (22)$$

经过上述处理后, 得到每一帧图像中所有粒子的分布状态, 粒子群密度越大运动目标存在可能性越大, 因此可以通过计算粒子群位置均值估计出目标位置信息, 计算公式如式(23):

$$X_{\text{target}} = \sum_{i=1}^N \omega_i X_i^i \quad (23)$$

4 实验分析

本节对基于粒子滤波的注意力模型及运动目标检测效果进行测试和分析. 在测试实验中, 将 B-U 运动注意力和 T-D 目标颜色注意力作为该模型的输入, 与其它相关方法分析对比目标注意力显著图, 同时评价对运动目标检测的准确性. 实验在 Dell Core 2.0GHz、1GB RAM 的计算机上完成, 编程环境为 Matlab2012.

4.1 实验设置

实验主要针对复杂运动场景测试运动目标检测的效果, 本文方法(基于粒子滤波的算法, Particle Filtering Based, PF)与文献[13](全局运动补偿和视觉注意力混合的目标分割, Globe Motion Compensation and Visual Attention Based, GMC-VAB)中的算法和文献[8]双向注意力融合的算法, Two-way Motion Attention Fusion Based, TWAFB)作对比. GMC-VAB 模型的相关参数设定如表1所示; TWAFB 模型的融合权值系数和偏置系数设定如表2所示; 本文方法中, 时间注意力 S_1 和空间注意力 S_2 的融合系数 α 与 β , B-U 注意力概率似然函数中标准差 σ_{B-U} 与 T-D 注意力概率似然函数中标准差 σ_{T-D} 的设定如表3所示.

4.2 注意力显著图与运动目标检测

在实验数据方面, 选取6个不同具有运动较为强烈的视频序列作为测试数据. 6种测试序列具有不同的运动特点, “飞行器”序列: 待测目标选定为飞行员, 目标在空中高速滑动, 背景为地面远景物体, 纹理复杂, 飞行器向图像上方运动, 镜头随飞行器移动; “斗牛士”序列: 待测目标选定为牛, 目标不规则的激烈运动, 背景为围栏和土地等, 纹理有一定起伏变化, 镜头随目标快速移动; “帆船”序列: 待测目标选定为较大白色帆船, 目标向右方做较缓慢平移运动, 且存在部分遮挡干扰, 远景纹理较弱, 远景较复杂; “马”序列: 待测目标选定为马, 目标激烈运动, 背景为树木和草地, 纹理变化平坦, 马向图像左侧奔跑, 镜头快速向左侧移动; “蜂鸟”序列: 待测目标选定为蜂鸟, 目标激烈运动, 背景有较强的明暗变化, 蜂鸟的头部在不停啃食桌面食物, 翅膀高速煽动, 镜头随蜂鸟位置变化轻微移动. “赛跑”序列: 待

测目标选定为运动员,目标与镜头都做剧烈高速运动,图像出现局部模糊和幻影.图5显示了测试结果,其中第一列为视频序列的原始帧;第二列为显著二值图;第三列为 Itti 算法结果,主要反映颜色、亮度、方向等静态特征形成的 B-U 注意力显著图;第四列为以颜色特征为基础的 T-D 注意力显著图;第五列为 MA 算法结果,要反映其 B-U 运动模型计算得到的运动注意力显著图;第六列为 GMC-VAB 算法结果,主要反映全局运动补偿后的运动注意力与静态注意力加权融合形成注意力显著图;第七列为 TWAFB 算法结果,主要反映 B-U 注意力与 T-D 注意力偏置加权融合形成注意力显著图;第八列为本文算法结果;第九列为本文算法的目标检测定位结果.上述测试结果显示 PF 算法总体效果优于 MA 算法、GMC-VAB 算法、TWAFB 算法. MA 算法虽然进行了双向注意力融合,但是受到运动估计精度的影响,导致运动注意力计算不准确,当运动相对简单时,可获得较好的效果,如“飞行器”和“蜂鸟”序列,而当运动强烈时,运动注意力比较杂乱不能有效反应运动目标

显著性,如“斗牛士”、“帆船”、“马”和“赛跑”序列; GMC-VAB 算法融合了全局运动补偿和静态注意力,但通常由于全局运动估计存在误差,导致目标显著性效果不佳;而且 GMC-VAB 算法只融合了 B-U 注意力特征,对全局中所有可能引起注意力的区域全部进行标注,无法实现目标的检测定位,如“马”,“赛跑”,“帆船”,“飞行器”,“蜂鸟”,“斗牛士”等序列, TWAFB 算法采用 T-D 注意力作为指导,得到结果相对较好,但是在角点特征不明显的序列中,如“帆船”、“蜂鸟”序列, TWAFB 的噪声较大,注意力显著焦点位置空间不连续,有时存在较大误差,实验结果显示, PF 算法注意力显著图比 MA 算法、GMC-VAB 算法、TWAFB 算法的注意力图,由于采用合理的处理方法提高了运动估计的精度,加入颜色特征作为指导,背景噪声干扰得到大幅度削弱,还有在粒子滤波过程,粒子分布状态随运动状态和目标位置而改变,粒子最终收敛聚集于运动目标区域,使得注意力准确性大大提高.

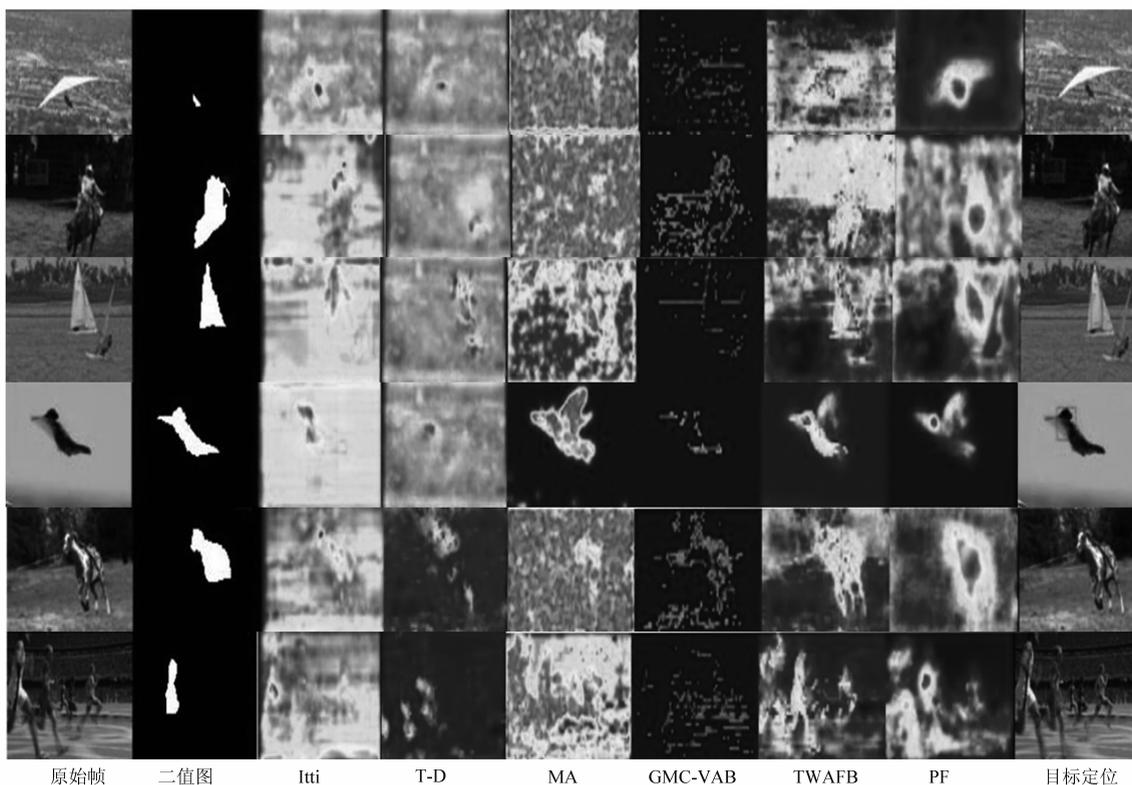


图5 六种测试序列检测结果

表 1 GMC-VAB 模型的参数

	偏置控制 δ	阈值 η
GMC-VAB	0.9	0.8

表 2 TWAFB 模型的参数

	T-D 权值 α	B-U 权值 $(1 - \alpha)$	偏置系数 β
TWAFB	0.5	0.5	1.2

表 3 PF 模型的参数

	融合系数 α	融合系数 β	标准差 σ_{B-U}	标准差 σ_{T-D}
PF	0.5	0.5	0.3	0.2

4.3 ROC 性能指标

本文采用特征曲线(ROC)来评价本文模型所得到的显著图的预测注意力目标的性能. 显著图中每个像素点被设定为一个二元分类器的样本, 像素点值大于某一阈值可以判为是注意力焦点(阳性样本), 否则判为非注意力焦点(阴性样本), 以图像帧的真实显著二值图作为标准(ground truth), 通过改变阈值大小, 得出系列真阳性率(R_{TP})和伪阳性率(R_{FP})对应的值, 分别以 R_{FP} 和 R_{TP} 为横轴与纵轴可以得到 ROC 曲线. R_{TP} 和

R_{FP} 由式(26)和(27)计算得到:

$$R_{TP} = M_{TP}/n_P \quad (26)$$

$$R_{FP} = M_{FP}/n_N \quad (27)$$

其中, n_P 和 n_N 分别为阳性样本和阴性样本个数, M_{TP} 和 M_{FP} 分别为判断为真阳性样本和伪阳性样本的个数.

图 6 显示了 6 个测试序列的 ROC 曲线结果. 曲线越靠近左上方, 运动目标显著性的准确度越高. 由 ROC 曲线可知本文注意力模型的性能要优于其它模型; TWAFB 模型加性融合 T-D 和 B-U 注意力, 其效果比 MA 模型(B-U 注意力)要好; GMA-VAB 由于全局运动补偿的缺陷, 导致效果不理想; Itti 模型属于静态特征(颜色、亮度和方向)产生的注意力, 所以曲线性能较差.

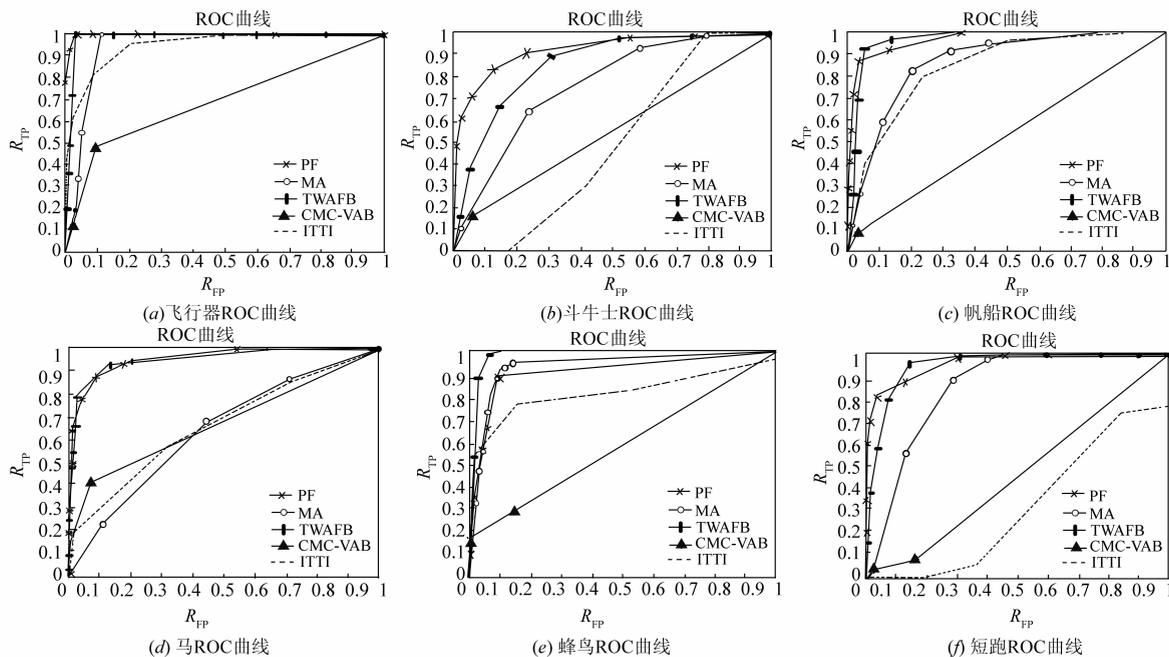


图6 注意力模型的ROC曲线

5 结论

本文依据贝叶斯融合原理, 构建了基于粒子滤波的视觉双向融合注意力模型, 并以运动注意力和目标颜色注意力分别为 B-U 和 T-D 注意力输入, 通过粒子权值计算、重采样等处理改变粒子分布状态, 形成注意力显著图, 并最终确定目标位置; 测试结果表明, 本文方法可获得较准确的目标注意力显著图, 对相关研究领域具有较高的学术和应用价值.

参考文献

- [1] Itti L, Koch C. Computational modeling of visual attention [J]. Nature Reviews Neuroscience, 2001, 2(3): 193 - 203.
- [2] Olivier L M, Patrick L C. A coherent computational approach to model bottom-up visual attention [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5): 802 - 817.
- [3] Hou X D, Zhang L Q. Saliency Detection: A spectral residual approach [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Minneapolis: IEEE, 2007. 1 - 8.
- [4] Rapantzikos K, Tsapatsoulis N, Avrithis Y, et al. Bottom-up spatiotemporal visual attention model for video analysis [J]. Image Processing, 2007, 1(2): 237 - 248.
- [5] Yan Q, Xu L, Shi J P, et al. Hierarchical saliency detection [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE, 2013. 1155 - 1162.
- [6] Kang W L, Hilary B, Feng J F. Cue-guided search: a com-

- putational model of selective attention [J]. IEEE Transactions on Neural Networks, 2005, 16 (4): 910 - 924.
- [7] Zhang L Y, Tim K Marks, Matthew H Tong, et al. SUN: A bayesian framework for saliency using natural statistics [J]. Journal of Vision, 2008, 7(8): 1 - 20.
- [8] Sang W B, Bumhwi K, Minho L. Top-down visual selective attention model combined with bottom-up saliency map for incremental perception [A]. IEEE International Conference on Neural Networks [C]. Barcelona; IEEE, 2010. 1-8.
- [9] Fang Y M, Lin W S, Lau C T, et al. A visual attention model combining top-down and bottom-up mechanisms for salient object detection [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Singapore; IEEE, 2011. 1293 - 1296.
- [10] Borji A. Boosting bottom-up and top-down visual features for saliency estimation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence; IEEE, 2012. 438 - 445.
- [11] Yu Y L, George K I Mann, et al. An object-based visual attention model for robotic applications [J]. IEEE Transactions on System, 2010, 40(5): 1398 - 1412.
- [12] 郎丛妍, 须德, 李冰. 一种基于动态贝叶斯网的视觉注意力是别方法 [J]. 电子学报, 2011, 39(3A): 140 - 146. Lang Cong-yan, Xu De, Li Bing. Dynamic Bayesian network based visual focus of attention recognition [J]. Acta Electronica Sinica, 2011, 39 (3A): 140 - 146. (in Chinese)
- [13] Ma Y F, Zhang H J. A model of motion attention for video skimming [A]. 2002 International Conference on Image Processing [C]. New York; IEEE, 2002. 129 - 132.
- [14] Zhou C. A visual attention model for dynamic scenes based on motion features [A]. IEEE International Conference on Control Automation Robotics & Vision [C]. Guangzhou; IEEE, 2012. 1396 - 1401.
- [15] Han J W. Object segmentation from consumer videos: a unified framework based on visual attention [J]. IEEE Transactions on Consumer Electronics, 2009, 55(3): 1597 - 1605.
- [16] Gordon N, Salmond D J, Smith A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation [J]. IEEE Proceedings F: Radar and Signal Processing, 1993, 140(2): 107 - 113.

作者简介



刘 龙 男, 1976 年生, 博士, 副教授, 研究方向为智能信息技术与嵌入式系统.

E-mail: Liulong@xaut.edu.cn

樊波阳 男, 1990 年生, 硕士研究生, 研究领域为智能信息技术与嵌入式系统.

刘金星 男, 1989 年生, 硕士研究生, 研究领域为目标检测与跟踪.