

# 基于保边滤波的显著目标快速分割方法

张雷<sup>1</sup> 李成龙<sup>1</sup> 涂铮铮<sup>1</sup> 汤进<sup>1,2</sup>

(1. 安徽大学计算机科学与技术学院, 合肥, 230601; 2. 安徽省工业图像处理与分析重点实验室, 合肥, 230039)

**摘要:** 在视频中自动发掘目标并对其进行精确分割是一个非常具有挑战性的计算机视觉问题。本文提出了一种基于保边滤波的显著目标快速分割方法。首先, 通过融合外观特征与运动特征, 将视频中的显著目标发掘转为能量函数最小化问题进行求解。其次, 为了更精确地进行分割目标, 融合外观的高斯混合外观模型(Gaussian mixture model, GMM)、位置先验以及时空平滑约束构建马尔科夫随机场(Markov random field, MRF)模型, 并使用图割算法进行求解。本文提出的基于保边滤波的显著目标快速分割方法, 在牺牲较少的精度下, 极大地提高了分割效率。最后在两个数据集上进行了对比实验, 实验结果表明, 本文算法的分割精度超过了其他5种目标分割方法, 且加速算法在损失少量精度的情况下提高了2倍分割效率。

**关键词:** 显著目标发掘; MRF模型; 保边滤波; 快速目标分割

**中图分类号:** TP391      **文献标志码:** A

## Fast Salient Object Segmentation Method Based on Edge-Preserving Filtering

Zhang Lei<sup>1</sup>, Li Chenglong<sup>1</sup>, Tu Zhengzheng<sup>1</sup>, Tang Jin<sup>1,2</sup>

(1. Department of Computer Science and Technology, Anhui University, Hefei, 230601, China; 2. Key Lab of Industrial Image Processing & Analysis of Anhui Province, Hefei, 230039, China)

**Abstract:** How to automatically discover salient objects in video and further perform accurate object segmentation is a challenging problem in computer vision. Here, fast salient object segmentation method based on edge-preserving filtering is proposed. Firstly, the salient object discovery is formulated as an energy minimization problem, which fuses the appearance and motion features. Then, a Markov random field (MRF) model, integrating the Gaussian mixture model (GMM) of appearance, the location prior, and the spatial-temporal smoothness, is constructed for accurate segmentation, and is efficiently optimized by graph cut. Moreover, an edge-preserving-based method is presented to improve the segmentation efficiency with a little loss of accuracy. Finally, extensive experiments on two datasets suggest that the proposed method performance is better than that of other five methods, and the accelerated version can speed up to 2 times of the original one.

**Key words:** salient object discovery; MRF model; edge-preserving filtering; fast object segmentation

**基金项目:** 国家自然科学基金(61472002)资助项目; 安徽省高等学校省级自然科学研究项目重点项目(KJ2014A015)资助项目; 安徽省高等学校自然科学研究重点项目(KJ2015A110)资助项目; 安徽省科技厅自然科学基金面上项目(1308085MF97)资助项目。

**收稿日期:** 2015-09-24; **修订日期:** 2015-11-24

## 引言

视频中的目标分割是计算机视觉中的一个基础底层问题,可用于运动识别、三维重建、视频检索和视频摘要等。由于无约束的视频中存在背景复杂、光照变化、目标和背景运动剧烈等因素,使得如何自动发掘目标,并对其进行准确地时空分割具有非常大的挑战性。近年来,一些研究学者根据目标的一些外观或运动特征实现目标的自动发掘,并进一步精确分割<sup>[1-8,10-11]</sup>。

(1)基于图像显著性的目标分割方法。该方法是在图像中利用目标显著性进行初始化,进而精确分割。这些方法可以看作是视频中目标分割的特例,即初始化和分割均不考虑时序关系。Cheng等<sup>[3]</sup>通过全局的颜色特征对比计算目标的显著性,然后阈值化作为目标的初始分割,进而使用GrabCut进行精确分割。该方法具有较快的处理速度,但鲁棒性不高。Achanta等<sup>[4]</sup>以像素点为中心的小窗口内Lab颜色特征与附近其他窗口进行局部对比得到显著性计算结果,并结合在多尺度上的显著性值得到最终的图像显著性检测结果,最后在过分割的基础上,将平均显著性值高于阈值的区域分割出来,以此作为目标的分割结果,此方法容易导致丢失大部分的目标区域。以上这些方法都是用图像显著性结果作为目标的初始化,实现自动的目标分割。而视频中的场景更为复杂,仅仅利用视频帧自身的信息不能有效地进行目标分割,所以方法存在较大的局限性。

(2)基于视频显著性的目标分割方法。Fukuchi等<sup>[5]</sup>结合当前帧的图像显著性检测结果与前一帧的分割结果,并用卡尔曼滤波得到当前帧的初始化目标,最后在马尔科夫随机场(Markov random field, MRF)中进行目标分割的求解,在图像处理单元(Graphics processing unit, GPU)加速下使得此方法具有较高的分割效率,但此方法在构建马尔科夫随机场时,并仅仅用了目标的外观特征,没有考虑到目标在视频中存在的位置先验信息,从而导致对一些由运动到静止的目标出现漏分割的情况。Li等<sup>[6]</sup>首先通过计算得到视频帧的图像与运动显著性信息,并由此信息得到的形状特征、前景与背景颜色特征以及显著性特征,并结合时空一致性约束构建条件随机场(Condition random field, CRF)模型,最后用图割的方法求解得到目标分割的结果,但此方法在显著性计算时不能均匀的凸显显著目标,且具有较高的时间复杂度,导致其应用受到较大的限制。Wang等<sup>[7]</sup>先计算每一视频帧的空间边缘特征与运动边缘特征,再对这两种特征计算测地距离得到显著性结果,并在此基础上通过计算与估计背景区域的测地距离得到较准确的显著性检测结果,最后结合全局的外观模型、位置先验以及时空平滑建立能量函数进行目标分割,是一种离线的目标分割方法。

(3)其他的视频目标自动分割方法。Lee等<sup>[8]</sup>首先把所有视频帧过分割成超像素<sup>[9]</sup>,并依据外观与运动特征差异计算超像素的显著值。然后对超像素进行聚类,利用超像素的显著值对类进行排序,把排名最高的类作为目标的初始化。最后以像素点为结点,结合目标和背景的高斯混合模型(Gaussian mixture model, GMM),以及局部形状匹配位置先验,构建MRF模型进行目标分割。在此基础上,Zhang等<sup>[10]</sup>构造了一个分层的有向无环图(Directed acyclic graph, DAG)来选取最终的目标区域,并利用光流对目标区域进行扩张。但Lee和Zhang等的分割方法只适合单目标的分割情况,且时间复杂度较高。Papazoglou等<sup>[11]</sup>通过目标的运动特征度量来初始化目标,然后结合位置先验以及时空平滑约束建立MRF模型快速分割出目标。由于该方法只采用了目标运动特征度量,在静止或者运动较小的目标分割中往往会失败。值得注意的是,文献<sup>[7-9,11]</sup>的目标分割方法都利用了视频的全局信息,导致这些方法都不能处理任意长的视频,具有较大的局限性。

本文针对以上目标分割中存在的问题,提出了一种基于保边滤波的显著目标快速分割方法。该方法能够在线自动地发掘视频中的显著目标,并在保边滤波的作用下保持目标的边界信息,得到较准确的目标初始化,在此基础上对目标进行快速分割,且该方法具有较高的分割精度。本文实验也验证了该方法的有效性,分割精度优于现有的其他目标分割方法,且具有较高的分割效率。

## 1 基于保边滤波的显著目标发掘和分割

本文提出基于保边滤波的显著目标快速分割方法,其算法流程图如图 1 所示。该方法首先在梯度驱动下对视频帧进行降采样,保留最大的梯度信息得到低分辨率的视频帧,并提取此低分辨率视频帧的外观特征与运动特征,然后在能量最小化的框架下融合这两种不同的特征,得到视频显著性检测结果,并均匀的凸显出显著目标。在此基础上进行阈值化发掘显著目标,并估计出目标和背景的外观模型,即高斯混合模型,再结合由上一帧传递到该帧的目标位置先验以及时空的平滑约束建立 MRF 模型,并使用图割算法进行求解得到目标精确的分割结果,最后还原到原分辨率大小,并用保边的局部多点交叉滤波(Cross-based local multipoint filtering, CLMF)算法进行处理得到的目标分割最终结果。

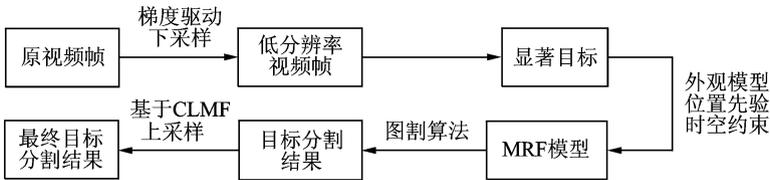


图 1 基于保边滤波的显著目标快速分割算法流程图

Fig. 1 Flow chart of fast salient object segmentation based on edge-preserving filtering

### 1.1 融合外观和运动特征的显著目标发掘

本文在能量最小化的框架下,通过融合外观特征和运动特征得到视频显著性计算结果,并用固定的阈值处理发掘显著目标。常用的图像外观特征包括颜色、亮度、纹理和边缘等,为了简单有效,本文使用具有互补作用的颜色与结构特征。其中,颜色特征采用 Lab 空间来表示,结构特征融合梯度方向和大小来表示,即在以像素点为中心的小窗口内,分别计算此窗口内像素的梯度方向和梯度大小的直方图,然后联接两个直方图作为该像素点的结构特征表示<sup>[12]</sup>。对于运动特征,使用光流<sup>[13]</sup>幅值大小来表示。

#### 1.1.1 视频显著性的公式形式化表示

本文在能量最小化框架下求解视频显著性问题。为了提取显著的目标以及一致的凸显目标区域,定义能量函数为

$$E = \sum_{p \in I} \{ \| H(p) - f_a(p) \|_2^2 + \lambda \| H(p) - f_m(p) \|_2^2 + C(H(p)) \} \quad (1)$$

式中:  $H(p) \in \{0, 1, \dots, h-1\}$  表示为像素点  $p$  分配的显著性等级,  $h$  表示最高显著性等级;  $f_a(p)$  和  $f_m(p)$  分别表示像素点  $p$  归一化后的外观特征和运动特征;  $\lambda$  表示权重参数;  $C(H(p))$  表示平滑项,定义见式(2)。由于在求解能量函数  $E$  时,平滑项作为权重进行计算,故并不需要在式(1)中加入权值。

$$C(H(p)) = \sum_{k \in \Omega_p} \omega_{pk} \| H(p) - H(k) \|_2^2 \quad (2)$$

式中:  $\Omega_p$  表示由像素点  $p$  生成的不规则自适应大小的联通区域,即在一定的臂长  $r$  范围内,根据每一个像素点  $p$  与此范围内其他像素点颜色距离小于阈值  $T_1$ ,生成联通的自适应大小窗口  $\Omega_p$ 。并用窗口内其他像素点  $k$  对其进行加权,权值大小为  $\omega_{pk}$ ,此权值由  $\Omega_k$  所包含的像素点确定,详见式(3)。因为对  $p$  加权由  $p$  生成的自适应窗口中其他像素点所决定,所以具有很好的保边作用。

$$\omega_{pk} = \frac{|\Omega_k|}{\sum_{k \in \Omega_p} |\Omega_k|} \quad (3)$$

式中:  $|\cdot|$  表示由像素点  $k$  生成的不规则联通区域  $\Omega_k$  所包含的像素点数。

### 1.1.2 基于 CLMF 的视频显著性求解

对能量函数  $E$  进行最小化求解, 本文采用代价体滤波技术<sup>[14]</sup>, 具体过程包含以下 3 个步骤, 输出的  $H(p)$  为像素点  $p$  的最终显著等级。

(1) 构建代价体

$$V_{p,H} = \| H(p) - f_a(p) \|_2^2 + \lambda \| H(p) - f_m(p) \|_2^2 \quad (4)$$

(2) 对代价体进行 CLMF 保边滤波

$$V'_{p,H} = \sum_{k \in \Omega_p} \omega_{pk} V_{k,H} \quad (5)$$

(3) 选择最优值

$$\hat{H}(p) = \operatorname{argmin}_H V'_{p,H} \quad (6)$$

图 2 为有 CLMF 作用下的视频显著性计算结果, 与没有 CLMF 的作用下的视频显著性计算结果的比较图。图 2 中依次给出了原视频帧、手工标注结果、在 CLMF 作用下的视频显著性计算结果以及没有 CLMF 作用下的视频显著性计算结果。从图中可以看出, CLMF 保边滤波很好地保持了目标边界信息, 均匀的凸显目标区域, 且能有效地去除噪声的干扰。



(a) 原视频帧

(a) Original video frame



(b) 手工标注结果

(b) Manually labeling result



(c) 在 CLMF 作用下的显著性计算结果

(c) Saliency calculation result with CLMF



(d) 没有 CLMF 作用下的显著性计算结果

(d) Saliency calculation result without CLMF

图 2 CLMF 在视频显著计算中的作用比较图

Fig. 2 Effect comparison images of CLMF in video saliency calculation

### 1.1.3 显著目标提取

本文通过在能量最小化框架下计算视频显著性, 得到显著性结果较好的均匀凸显出目标区域, 且很好地保持目标的边界信息, 并利用简单的阈值  $T_2$  分割得到显著目标, 在此基础上用 MRF 模型对目标进行精确的分割。

## 1.2 基于 MRF 模型的目标精确分割

对于进一步的精确分割, 本文以超像素<sup>[15]</sup>为处理单位, 首先通过初始分割结果和超像素的重叠比确定目标超像素。然后以超像素为结点构建 MRF 模型, 能量函数定义如式(7)所示, 并使用 Graph Cut<sup>[16]</sup>进行求解得到目标精确的分割结果。

$$\mathbf{t}^* = \operatorname{argmin}_{\mathbf{t}} E(\mathbf{t}) = \operatorname{argmin}_{\mathbf{t}} \left( \sum_{t=1}^T A_t^i(A_i^{-1}, S_i) + \alpha_1 \sum_{t=1}^T L_t^i(I_i) + \alpha_2 \sum_{(i,j) \in \xi} V_{ij}^t(I_i, I_j) + \alpha_3 \sum_{(i,j) \in \xi} W_{ij}^t(I_i, I_j^{t+1}) \right) \quad (7)$$

式中:  $\mathbf{t} \in \{I_i\}_{t,i}$  表示所有超像素的标签集;  $A_t^i$  表示第  $t$  帧  $i$  个超像素的 GMM 外观模型;  $S_i$  表示  $t$  帧  $i$  个超像素的目标初始化结果;  $I_i \in \{0, 1\}$  表示  $t$  帧第  $i$  个超像素标签(0, 1 分别表示背景和目標);  $L_t^i$  表示  $t$  帧第  $i$  个超像素的位置先验, 它由  $t-1$  帧对应的超像素经过光流得到;  $\xi_t$  表示  $t$  帧中所有的超像素在空间上重叠边界的集合;  $\xi_t$  表示  $t-1$  帧和  $t$  帧中所有邻接的超像素边界的集合(对于  $t-1$  帧的超像素  $s_{t-1}$  和  $t$  帧的超像素  $s_t$ , 若  $s_{t-1}$  中的像素点经过光流运动后至少有一个像素点和  $s_t$  的像素点重叠, 则称  $s_{t-1}$  和  $s_t$  在时序上是邻接的);  $\alpha_1, \alpha_2, \alpha_3$  表示权衡因子;  $V_{ij}^t$  为空间平滑项, 表示超像素  $i, j$  在空间的平滑约束;  $W_{ij}^t$  为时序平滑项, 表示超像素  $i, j$  在时序上的平滑约束, 具体参看文献[11]。

### 1.3 基于梯度驱动以及 CLMF 保边滤波的加速算法

为了降低时间复杂度,满足视频数据的时间处理需求,本文提出一种基于梯度驱动以及保边滤波的加速算法,在损失少量精度的基础上极大提高了分割速度。首先使用  $m \times m$  的窗口对视频帧进行降采样,使用该窗口内梯度最大的像素点作为低分辨率视频帧对应位置的像素值;然后在此低分辨率视频上使用 1.1 节和 1.2 节的方法进行显著目标的分割;最后将分割出的结果升采样到原分辨率大小,并在原视频帧的引导下使用 CLMF 保边滤波进行处理,由此保证分割的精度。图 3 马的视频给出了梯度驱动降采样及 CLMF 保边滤波升采样的比较图,从图中可以看出,本文在梯度驱动降采样的基础上得到较好的分割结果,并在原视频帧的引导下进行保边滤波处理是有效的,较好地保持了目标的边界信息。

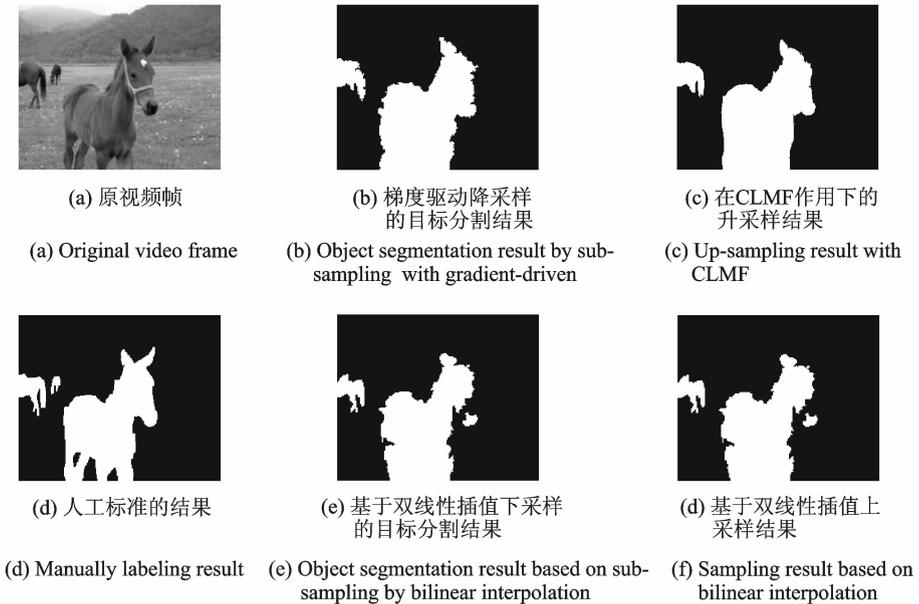


图 3 梯度驱动降采样及 CLMF 保边滤波升采样比较图

Fig. 3 Comparison between sub-sampling result with gradient-driven and up-sampling result with CLMF edge-preserving filtering

算法的具体流程如下。

输入:视频帧。

(1)梯度驱动的降采样。对视频帧用  $m \times m$  的窗口进行降采样,保留最大的梯度信息。

(2)显著目标发掘。在能量最小化的框架下,融合外观特征与运动特性进行视频中显著目标的发掘,具体细节见 1.1 节。

(3)GMM 估计。用当前帧显著目标的检测结果优化上一帧的 GMM 作为当前帧外观 GMM。

(4)目标分割。根据当前帧的外观 GMM 和位置先验以及时空的平滑约束求解式(8),得到目标分割结果,并输出当前帧的外观模型。

(5)在 CLMF 作用下的升采样。将目标分割结果升采样到原分辨率大小,并对此进行保边滤波得到原分辨率的目标分割结果。

输出:目标分割结果。

## 2 实验结果与分析

本文在两个数据集上与 5 种不同的目标分割方法进行比较,通过对实验结果的对比分析,验证了本

文方法的有效性,其分割精度在大部分情况下优于其他的目标分割算法,并取得了较高的分割效率。

## 2.1 实验设置

实验中的两个数据集为 Fukuchi<sup>[5]</sup> 公共数据集与自己收集的视频集。其中 Fukuchi 公共数据集包含 10 个不同的视频,分别为飞机、鸟、狐狸、滑雪 1、滑雪 2、马、猫、黄鹌、犀牛和向日葵,并给出了手工标注的分割结果。为了进一步验证本文算法的有效性以及实用性,本文也收集了 6 个不同类型的视频进行实验,其中选取了 YouTube-Objects 公共数据上的飞机、马和摩托车这 3 个视频中的一个镜头,以及自己拍摄的监控场景的行人、两个行人、两人交谈这 3 个视频,并都进行了人工标注。为了实验比较的客观性,在选取这些视频时,考虑到文献[11]只对运动目标才能够进行有效分割以及文献[10]只适用于单目标分割的情况,本文选取视频中的飞机、摩托车和两个行人都有较大的运动,而飞机、摩托车和马的视频中都是单一的目标。总体上,这两个数据集考虑到了目标的个数、目标运动的快慢、镜头远近以及摄像机的运动,能够较全面地评价分割算法的性能。

本文实验中的比较方法包括现有较流行的目标分割方法,其中包括文献[6]、文献[10]、文献[11]、VIBE<sup>[17]</sup> 和 RCC<sup>[3]</sup> 的目标分割算法,文献[6,10,11]算法是近年较好的目标分割算法,VIBE 是经典的目标检测算法,而 RCC 是基于图像显著性的目标分割算法。在实验中设置参数如下: $\lambda, h, \alpha_1, \alpha_2, \alpha_3, T_1$  和  $r$  分别设置为 0.6, 24, 1, 5 000, 4 000, 25 和 9,视频显著性检测结果归一化到[0,1]之间,阈值  $T_2$  设置为 0.4,  $m \times m$  大小的窗口设置为  $2 \times 2$ ,并假设前景与背景分别由 5 和 8 个单高斯组成,且在实验中这些参数都保持不变。

## 2.2 实验结果与分析

目前大部分目标分割算法的精度比较,都是在人工标注的基础上统计错误率,式(8)中 error 表示视频的平均错误率,  $p, p' \in \{0, 1\}$  (0, 1 分别表示背景和目标),  $F$  和  $GT$  分别为分割结果和人工标注的结果,  $XOR(\cdot)$  为异或操作,  $N$  为视频帧分辨率大小,  $FS$  表示此视频所含的视频帧数。

$$\text{error} = \frac{1}{N \cdot FS} \sum_{p \in F, p' \in GT} \text{XOR}(p, p') \quad (8)$$

### 2.2.1 定量分析

表 1 中给出了在 Fukuchi 公共数据集上的实验精度比较结果,其中第 1 列是此数据集中所有视频类别,括号内给出了视频的总帧数。第 2, 3, 4, 5, 6, 7 列分别表示本文算法、文献[10]、文献[11]、VIBE、RCC 和文献[6]的平均错误率,倒数第 1, 2 行分别表示本文算法相对于其他算法提高精度的百分比,以及每种算法在整个数据集上的平均错误率。从表中的平均错误率与提高精度百分比可以看出,本文方法的分割精度优于其他 5 种算法的分割精度,且相对于 VIBE 算法提高了 84.1% 的精度。RCC 算法在此数据集中表现出较好的分割精度,但本文算法在此基础上仍提高了 16.6% 的精度,且 RCC 算法只针对静态图像进行目标分割,没有充分利用视频的时序信息,具有一定的局限性,这从图 4 的精度比较曲线中可以看出。

图 4 给出了在本文收集的 6 个视频的分割精度比较曲线,其中不同的颜色代表不同的分割算法,横坐标表示视频

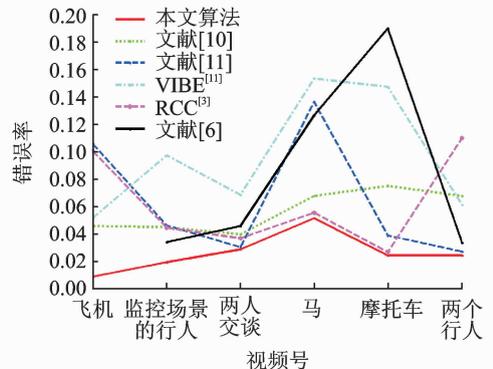


图 4 在收集的视频数据中每个视频的平均错误率比较折线图

Fig. 4 Average inaccuracy rate comparison line chart for every video in collected video datasets

表1 在 Fukuchi 公共数据集上分割算法的精度比较结果(其中黑色加粗为最高的分割精度)

Tab. 1 Accuracy comparison of different segmentation algorithms on Fukuchi public dataset (The black-bordered is the best segmentation accuracy)

算法	本文算法	文献[10]	文献[11]	VIBE <sup>[17]</sup>	RCC <sup>[3]</sup>	文献[6]
飞机(57)	0.023 5	0.035 2	<b>0.014 8</b>	0.063 2	0.046 4	0.059 8
红鸟(69)	0.024 3	0.057 9	0.112 8	0.111 7	0.017 5	<b>0.016 8</b>
狐狸(86)	<b>0.007 1</b>	0.077 4	0.008 8	0.128 7	0.053 0	0.022 2
滑雪 1(48)	<b>0.014 7</b>	0.067 3	0.043 3	0.320 4	0.119 2	0.064 3
滑雪 2(68)	<b>0.008 9</b>	0.075 5	0.033 2	0.248 8	0.038 8	0.040 1
马(74)	<b>0.031 7</b>	0.111 1	0.033 5	0.235 7	0.075 8	0.124 2
猫(92)	0.017 5	0.106 9	<b>0.011 9</b>	0.148 5	0.0190	0.088 8
黄鹂(103)	0.146 6	0.113 6	0.612 9	0.318 2	<b>0.037 3</b>	0.158 7
犀牛(85)	<b>0.007 2</b>	0.122 2	0.008 5	0.223 7	0.007 3	0.076 8
向日葵(86)	0.050 6	0.140 8	0.058 1	0.290 9	<b>0.031 6</b>	0.054 1
平均错误率	<b>0.033 2</b>	0.090 8	0.093 8	0.209 0	0.039 8	0.070 6
精度提高百分比/%	—	63.4	64.6	84.1	16.6	53.0

号,并给出了对应视频的名称,纵坐标表示平均错误率。由于文献[6]在飞机的视频中具有较高的错误率0.667 8,为了更好地进行折线图的比较,故在折线图未加入文献[6]飞机视频的精度值。从图中可以看出,本文的分割精度在大部分视频上都要优于其他方法的分割结果。其中文献[10]虽然整体的分割效果较稳定,但分割精度不高,且复杂度较高,很难得到广泛的应用。文献[11]在飞机与马的单个大目标的视频分割中出现了漏分割的情况,从而导致精度较低。VIBE算法对背景有运动干扰的视频具有较差的分割结果,如马和摩托车的视频。RCC算法在飞机与两个行人的视频中存在较多的分割错误,而在摩托车上表现较好的分割结果,这是由于摩托车视频中的目标具有较好的图像显著性,从而使得RCC具有较高的分割精度。由于飞机视频中镜头远近的不断变化,对文献[17]中提取的运动特征影响较大,导致分割的精度较低,而本文算法在这6个视频中都具有较好的分割结果,超过了其他的目标分割算法。

本文实验所用的PC处理器为Intel(R) Core(TM) i7-4790K CPU 4.00 GHz。视频帧大小为640 \* 360时,分割一帧所需时间约为1.2 s,但在本文实验中先降采样到320 \* 180,其中降分辨率与升分辨率时间约为0.04 s/帧,计算光流所用时间为0.08 s/帧,视频显著性计算每帧需要约0.2 s,总共分割一帧所需时间约为0.6 s,提高了约2倍的分割效率。文献[11]处理1帧的时间大概是0.4 s(除去计算光流和过分割时间),RCC处理一帧所需时间为0.004 s。尽管VIBE能够实时的进行目标分割,但在很多情况下不能得到较好的分割结果。文献[10]不仅内存需求大,而且耗时巨大,仅每一帧产生可能的目标区域所需时间就超过80 s,文献[6]计算每帧的运动与图像显著性的时间超过60 s,文献[6,10]的分割时间远大于本文方法的分割时间。

### 2.2.2 定性分析

本文展示了4组不同的视频分割结果,如图5所示。前两组视频来自Fukuchi公共数据集中的滑雪1与滑雪2的视频,后两组视频为自己收集的视频数据。每组视频给出了其中3帧的分割结果,左上角标注的序号为视频的帧号。其中每一组视频的第1,2,3,4,5,6行分别表示本文算法、文献[10]、文献[11],VIBE<sup>[17]</sup>,RCC<sup>[3]</sup>和文献[6]的分割结果。可以看出,本文的分割结果优于其他5种方法的分割结果。

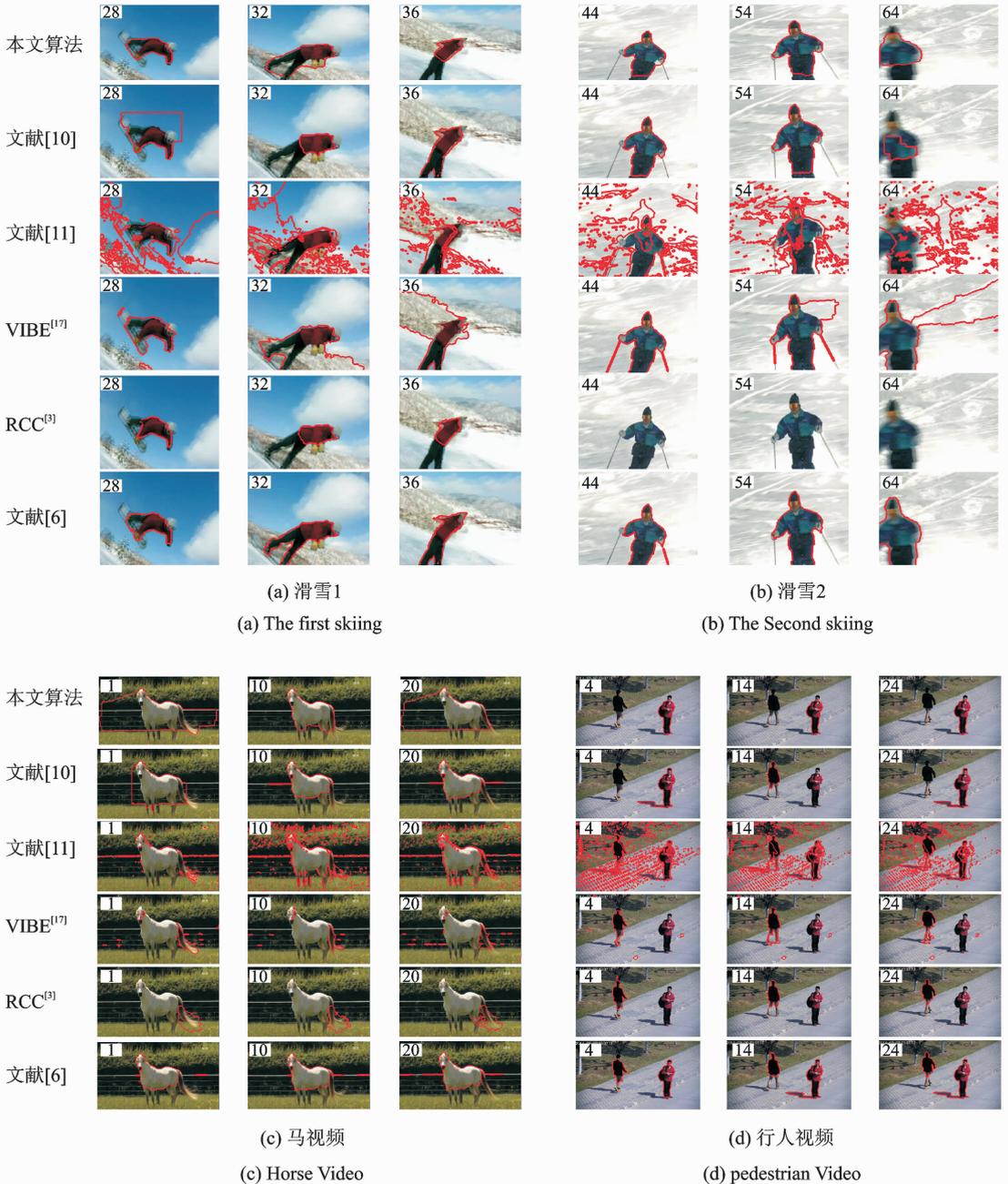


图 5 4 组视频中的部分帧分割结果比较图

Fig. 5 Some frame segmentation result comparison in four sets of videos

从 4 组视频第 2 行可以看出, 文献[10]对于单个大目标的马视频中, 只分割出了马的一部分, 且在第 1 组视频中没有分割出任何目标。从第 3 行看出, 文献[11]在目标运动较快时有较好的分割结果, 但对运动较小或者静止的目标不能进行有效的分割。从第 4 行看出, VIBE 算法在相机运动或背景运动干扰的情况下分割结果较差, 不能对目标进行有效分割。从第 5 行可以看出, RCC 算法对一些运动的目

标不能进行有效的分割,主要是由于 RCC 没有考虑到目标的运动信息,所以此算法对视频中的目标分割具有局限性。从第 6 行可以看出,文献[6]在马的视频中误分割出背景,对监控场景的行人存在漏分割的情况。从 4 组视频的第 1 行可以看出,本文算法基本克服了其他算法的缺点,对视频中的所有目标都具有较好的分割结果,但也存在一些问题,比如在第 4 组监控场景的行人视频中,将行人的影子也作为目标分割出来,后面将针对这一类问题进行研究,且由于本文是一种在线的目标分割算法,前几帧的分割受目标初始化的影响较大所导致前几帧的分割结果可能较差,但这也是在线目标分割都会遇到的问题。

### 3 结束语

本文提出了一种基于保边滤波的显著目标快速分割方法,该方法能够在无约束的场景下自动发掘不同类型、不同运动模式下的多个目标,并进行精确地分割。为了提高效率,使用一种基于梯度驱动以及保边滤波的加速算法,在保证精度的基础上,极大地提高了分割速度。实验验证了该方法的有效性,其精度在大部分情况下优于其他目标分割算法,且具有较高的分割效率。下一步的研究重点是融合目标的边界特征等联合地对目标建模,进一步地提高分割精度,以及提出相应的快速求解算法,降低处理复杂度,更好地服务于后续的其他视觉任务。

#### 参考文献:

- [1] 赵胜男,王文剑.一种快速均值漂移图像分割算法[J].数据采集与处理,2015,30(1):192-201.  
Zhao Shengnan,Wang Wenjian. Fast mean shift for image segmentation[J]. Journal of Data Acquisition and Processing, 2015, 30(1):192-201.
- [2] 唐利明,黄大荣,李可人.基于变分水平集的图像分割模型[J].数据采集与处理,2014,29(5):704-712.  
Tang Liming, Huang Darong, Li Keren. New model based on variational level set for image segmentation[J]. Journal of Data Acquisition and Processing, 2015, 29(5):192-201.
- [3] Cheng M, Mitra N J, Huang X, et al. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569-582.
- [4] Achanta R, Estrada F, Wils P, et al. Salient region detection and segmentation[C]// Computer Vision Systems. [S.l.]: Springer Berlin Heidelberg, 2008: 66-75.
- [5] Fukuchi K, Miyazato K, Kimura A, et al. Saliency-based video segmentation with graph cuts and sequentially updated priors [C]// Proceedings of IEEE International Conference on Multimedia and Expo. New York:[s. n.],2009: 638-641.
- [6] Li W T, Chang H S, Lien K C, et al. Exploring visual and motion saliency for automatic video object extraction[J]. IEEE Transactions on Image Processing, 2013, 22(7): 2600-2610.
- [7] Wang W, Shen J, Porikli F. Saliency-aware geodesic video object segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA:[s. n.],2015.
- [8] Lee Y J, Kim J, Grauman K. Key-segments for video object segmentation [C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain:IEEE,2011:1995-2002.
- [9] Endres I, Hoiem D. Category independent object proposals[C]//Proceedings of European Conference on Computer Vision. Hersonissos, Crete,Greece:[s. n. ],2010,635:575-588.
- [10] Zhang D, Javed O, Shah M. Video object segmentation through spatially accurate and temporally dense extraction of primary object regions[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA:IEEE, 2013: 628-635.
- [11] Papazoglou A, Ferrari V. Fast object segmentation in unconstrained video[C]// Proceedings of IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 1777-1784.
- [12] Wang Keze, Lin Liang, Lu Jiangbo, et al. PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence[J]. IEEE Transactions on Image Processing, 2015, 24(10): 3019-3033.

- [13] Bao L, Yang Q, Jin H. Fast edge-preserving patchmatch for large displacement optical flow[J]. *IEEE Transactions on Image Processing*, 2014, 23(12): 4996-5006.
- [14] Hosni A, Rhemann C, Bleyer M, et al. Fast cost-volume filtering for visual correspondence and beyond[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(2): 504-511.
- [15] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.
- [16] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(11): 1222-1239.
- [17] Barnich O, Van D M. Vibe: A universal background subtraction algorithm for video sequences[J]. *IEEE Transactions on Image Processing*, 2011, 20(6): 1709-1724.

#### 作者简介:



**张雷** (1990-), 男, 硕士研究生, 研究方向: 视频分析, E-mail: zhanglei\_edu@fox-mail.com。



**李成龙** (1988-), 男, 博士研究生, 研究方向: 计算机视觉和机器学习等。



**涂铮铮** (1982-), 女, 讲师, 研究方向: 图像处理与模式识别等。



**汤进** (1976-), 男, 教授、博士生导师, 研究方向: 计算机视觉、图像处理和模式识别等。