

基于局部图金字塔的不规则块匹配视频分割方法^{*}

彭小宁^{1,2} 杨明^{1,2} 邹北骥^{1,2} 陈姝^{1,2}

(中南大学信息科学与工程学院 长沙 410083)¹(湖南怀化学院计算机科学与技术系 怀化 418008)²

摘要 跳水、体操等视频中的运动员一般着装比较单一、肤色裸露,将这种视频帧图像用图的形式表示出来可以划分为少数的几个连通分量。根据这一特点,本文提出了一种基于局部图金字塔的不规则块匹配视频分割方法,该方法首先将图像以图的数据结构表示,采用假设检验的方法来合并相邻区域的图像块。在分割得到第一帧前景区域之后,根据帧间运动的连续性,运用不规则块匹配方法找到当前帧中的前景区域,匹配准则借鉴了聚类分析中变量相关性的相关系数方法。实验结果表明,这种方法能实时地、准确地分割出跳水、体操等视频中的运动员。

关键词 肤色裸露,局部图金字塔,不规则块匹配,全局运动,动态场景,感兴趣区域

Irregular Block Matching Video Segmentation Method Based on the Local Pyramid Graph

PENG Xiao-Ning^{1,2} YANG Ming^{1,2} ZOU Bei-Ji^{1,2} CHEN Shu^{1,2}

(School of Information Science and Engineering, Central South University, Changsha 410083)¹

(Department of Computer Science and Technology, HuaiHua College, HuaiHua 418008)²

Abstract The athletes in the diving and gymnastics sports videos always with simple dress and naked skin, if use the graph to represent the video frames we can get a few connected components. Based on this assertion, this paper proposes an irregular block matching video segmentation method based on the local pyramid graph. First, it represents the image by the graph data-structure, and then adopts the hypothesis test to merge the adjacent image block. After segmenting the first frame, according to the continuity of frame motion, it takes advantage of the irregular block matching method to find the foreground area in the current frame and take the correlation coefficient method of variable correlation in the clustering analysis as the matching criteria for reference. Experimental results show that this method can segment the athletes in the diving and gymnastics videos precisely in real time.

Keywords Naked skin, Localized pyramid, Irregular block matching, Global motion, Dynamic scene, Region of interest(ROI)

1 引言

为捕捉跳水、体操运动中的运动员所完成动作的完整动作序列,摄像机一般都需要移动摄像机镜头进行跟踪拍摄。这样由摄像机运动造成的视频背景的运动称作为全局运动,动态场景视频分割就是对存在全局运动的视频进行前景运动对象的提取。动态场景视频分割由于其不必对处理的视频作任何拍摄限制,它更具有一般性。因此,将视频中的运动对象从动态背景中分割出来有着十分重要的研究意义,它是基于视频的人体运动分析、基于对象的视频检索、面向对象的视频压缩码、基于视频的智能人机交互等基于内容的视频应用的基础。

目前,国内外动态场景视频对象分割的方法大概分为以下三类:一类是基于全局运动估计的方法^[1,2]。这类方法都是先通过摄像机运动模型来求得全局运动参数,然后进行运动补偿将视频序列中的若干帧背景对齐,以此来重建背景。当重建出背景之后,就可以按照静态场景的视频分割方法进行处理。效果比较好的静态场景视频分割方法主要有三帧差

分法^[3],运动检测法^[4],高斯背景模型法^[5]等。这类基于全局运动估计的方法分割效率和精确度主要取决于全局运动估计的效率和估计准确度。第二类是动态场景视频分割方法,即时空联合的方法^[6,7]。它们同时利用当前帧的纹理一致区域分割结果和相邻帧之间的变化检测结果。其中,最为典型的方法就是首先根据时域分割方法粗略地确定前景对象的大概区域,然后在对应的空域分割结果中精确地定位对象的边缘。例如:文^[6]提出了基于图像模型的时空联合分割,用马尔可夫随机场和贝叶斯决策来模型化运动矢量场、亮度分割场以及最终分割场之间的相互关系,实现了对运动背景序列的分割,但是由于亮度信息利用得不充分,分割的对象边缘精确度不高。第三类是交互式的视频分割方法^[8,9],这类方法牺牲分割的实时性以达到准确分割的目的。在交互式分割方法中,文^[8]结合了支持向量机和学习分类的方法,对部分静止背景的视频序列 Mom、Mother&Daughter 等能够获得较好的分割结果,但是对于运动背景序列的分割贡献很小,同时,由于需要大量的视频序列作为学习训练样本,分割过程繁琐,极大地影响了分割的效率,无法在实时系统中应用。

^{*}基金项目:国家自然科学基金(No. 60573079, No. 60673093);湖南省自然科学基金(No. 06JJ2065);怀化学院计算机应用技术重点学科项目资助课题;长江学者和创新团队发展计划资助。彭小宁 博士研究生,副教授,研究方向为图像处理与计算机视觉、数据库与信息安全技术;杨明 硕士研究生,研究方向为数字图像处理、视频分析;邹北骥 教授,博士,博士生导师,研究方向为图形学与 CAD 技术、数字图像处理;陈姝 博士研究生,研究方向为人体运动跟踪,智能监控,嵌入式系统应用。

由以上分析可知,视频对象的分割过程实际是一个语义特征提取与图像分割过程,而图像的分割本身就是计算机视觉和图像处理领域中很难解决的一个不确定性问题。现有的分割算法大多是针对某些具体问题的,至今尚无统一的、通用的理论和方法。我们针对具体的应用,根据跳水视频的特点,提出了一种基于局部图金字塔不规则块匹配的方法来进行视频分割。提出这种方法主要是基于跳水视频以下的特点:1)跳水运动员着装比较简单、肤色裸露;2)跳水运动员在垂直方向运动;3)摄像机一般只发生平移运动,基本上不会变焦;4)一段视频之内没有镜头的切换。充分利用这些特点,本文提出的方法能在保证准确分割的前提下,达到用户实时性的要求。

2 基于图模型的视频分割方法特点与原理

2.1 graph cuts 方法原理介绍

基于局部图金字塔的不规则块匹配视频分割方法,其本质上属于基于图模型的分割方法。为此,首先有必要了解一下 graph cuts 进行图像分割的原理,以便和本文提出的算法进行比较。

graph cuts 方法是图像以图的数据结构存储,运用图论中的一些相关知识来进行图像的分割。经典的 graph cuts 方法见文[10,11]。文[10]为规则切割法(Ncut),Ncut 方法采用自顶向下的方式进行图像分割。对图 $G(V, E)$ 的一个划分 V_1, V_2 , Ncut 的值直观地表示了这些顶点 V_i 和分割 (V_1, V_2) 连接的紧密程度。随着 Ncut 取值的减少,每一个分割内部的边的权值和与分割之外的边的权值和的比值增加。因此,如果用图表示图像,边的权值作为相似性度量的话, Ncut 方法给出了图像分割的度量方法,其可概括为越小的 Ncut 值,表示越好的分割效果。假设 I 表示 $n \times n$ 个像素的图像,为了把 I 表示成图 $G(V, E)$ 的形式,定义 V 为像素点的集合,连接每一个像素点的是它的四邻域。因此,图像被表示成网格图。边 (u, v) 的权值 $W(u, v)$ 表示 u, v 之间的相似性。相似性越高,权值越大。用权值函数 $W(u, v) = e^{-k|u-v|}$ 来表示。一旦把图像表示成图的形式,就可以迭代地应用 Ncut 方法分割图像。Ncut $V=(A, B)$ 定义成如下形式:

$$Ncut(A, B) = \frac{Cut(A, B)}{assoc(A, V)} + \frac{Cut(A, B)}{assoc(B, V)} \quad (1)$$

$$\text{其中, } Cut(A, B) = \sum_{v_i \in A, v_j \in B} W(i, j) \quad (2)$$

$$assoc(X, V) = \sum_{v_i \in X, v_j \in V} W(i, j) \quad (3)$$

因此, Ncut 图像分割问题就可以归约为找最小切割路径问题。

文[11]先构造一个图 $G(V, E)$, V 是相关顶点的集合,对于像素 P_i, P_j , 如果 $\|P_i - P_j\| < d$, 就为这两个像素之间建立一条边, d 是一个距离阈值, 权值函数定义为:

$$W((v_i, v_j)) = \|I(v_i) - I(v_j)\|, (v_i, v_j) \in E$$

其中, $I(v_i)$ 是像素点 v_i 的亮度值, 这里可以看出, $W(v_i, v_j)$ 是不相似性的度量, 权值越高, 越不相似, 这个定义刚好和 Ncut 相反。直观地, 这个算法把具有小权值的边聚类在一起, 为了聚类这些像素点, 定义了如下一些规则, 符合这些规则就进行聚类。如果 S 是 $G(V, E)$ 的一个分割, C 是 S 中的某个分量, $M(C, E)$ 是包含边集合 E 的 C 的最小生成树(MST)。算法使用如下的“分裂”和“合并”准则:

C 分量的内部差异:

$$Int(C) = \max_{e \in MST(C, E)} W(e) \quad (4)$$

分量 C_1, C_2 之间的差异:

$$Dif(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2} W((v_i, v_j)) \quad (5)$$

分量 C_1, C_2 的合并准则:

$$Dif(C_1, C_2) \leq Mint(C_1, C_2) \quad (6)$$

其中,

$$MInt(C_1, C_2) = \min \left(Int(C_1) + \frac{K}{\|C_1\|} Int(C_2) + \frac{K}{\|C_2\|} \right) \quad (7)$$

$\|C\|$ 为 C 分量的大小, K 是一常量。

纵观这两种图切割的方法, 可以统一到一点就是两者都要从图中寻找到最小切割路径, 每次以此路径作为图像的一个切割, 如此迭代地进行, 直到分割至迭代终止条件。但是, 文[12]已经证明图中的最大流路径就是最小切割路径, 并且求图中最大流路径的算法为 NP 难问题。

2.2 基于局部图金字塔的不规则块匹配视频分割方法 (LPIBM)

2.2.1 LPIBM 算法流程

基于局部图金字塔的不规则块匹配视频分割方法 (LPIBM), 先由人体肤色分布的先验知识确定大概的前景区域, 对背景比较单一的第一帧图像可以通过自适应阈值自动分割得到前景。如果第一帧图像背景纹理丰富, 场景复杂的话, 有必要对首帧进行交互式分割的方式来得到准确的前景。根据视频帧间的连续性, 由第一帧前景的位置, 可以确定第二帧的感兴趣区域 (ROI), 并以第一帧分割得到的前景为标准在后续帧中进行不规则块匹配。LPIBM 算法流程如图 1 所示。

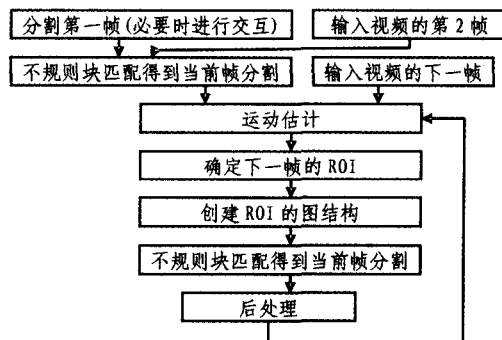


图 1 LPIBM 算法流程图

2.2.2 LPIBM 方法特点

首先, 空域方面, 针对跳水视频的特点: 跳水运动员着装比较简单、肤色裸露, 这就给出了前景对象即运动员的色彩和纹理信息, 利用这一信息, 可以进行对象轮廓的提取、对象纹理的分析以及色彩范围的限定, 以此来达到帧内分割的目的; 时域方面, 根据跳水视频的以下特点: 1) 摄像机一般只发生平移运动, 基本上不会变焦; 2) 跳水运动员在垂直方向运动; 3) 一段视频之内没有镜头的切换。首先来看第一个特点, 摄像机只发生平移运动说明在进行运动估计的时候可以忽略旋转和缩放参数, 对式(8)的 6 参数仿射运动模型, 可以简化模型中的运动参数, 得到新的简单的摄像机运动模型, 如式(9)所示。

$$\begin{cases} x = ax' + by' + e \\ y = cx' + dy' + f \end{cases} \quad (8)$$

其中, $p=(x, y)$ 为当前帧 I_k 的坐标, $p'=(x', y')$ 为相邻帧

I_k 中与 p 对应的点的坐标。 (a, b, c, d, e, f) 为全局运动参数, a, b, c, d 表示旋转和缩放, e, f 表示位移。

简化后的仿射运动模型可以表示为:

$$\begin{cases} x = x' + e \\ y = y' + f \end{cases} \quad (9)$$

x, y, x', y', e, f 表示的含义与式(8)相同。

第 2, 3 个特点说明了视频是一段运动集中的 Y 方向上的连续视频, 如果把视频分成帧来看的话, 相邻两帧之间的变化可以限定在一个范围之内, 根据我们的实验, 对于一般的跳水视频, 相邻两帧之间的变化不会超过 ± 12 个像素。利用这一信息, 在进行运动估计的时候, 采用最准确的块匹配算法——全搜索算法时, 可以限定搜索的窗口大小为: $(-12, -12) \times (12, 12)$, 这样可以极大地提高搜索的时间效率。

从 LPIBM 方法的以上两个特点来看, 我们在进行视频分割的时候既利用了视频序列图像之间的时域信息, 又利用了单帧图像内部的空域信息。所以, 基于局部图金字塔不规则块匹配视频分割算法是一种时空联合的方法。

2.3 图的创建与聚类

与 graph cuts 方法一样, 本文的方法在对图像处理之前, 也需要建立图像的图的数据结构, 如图 2 所示。

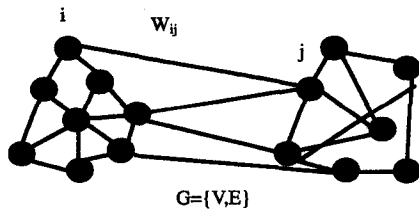


图 2 图像以图的存储形式

其中, V 为图的顶点集合, 对应着图像中的像素点; E 为连接顶点的边的集合, 边上的权值 W_{ij} 为相邻像素间的相似性。这样就把图像分割的问题转化为图的切分问题。

为了提高图创建的合理性和健壮性, 本文采用了以种子像素点为参考点的图像四邻域比较。具体做法为: 如果该像素点的四邻域与种子像素点的颜色距离小于给定的阈值 d , 则在该像素与其邻域像素间建立一条边, 边的权值 W_{ij} 为这两个像素点之间的颜色距离。如式(10)所示。

$$W_{ij} = \begin{cases} |I(V_i) - I(V_j)|, (V_i, V_j) \in E \\ \infty, \text{otherwise} \end{cases} \quad (10)$$

其中, $I(V_i)$ 为 P_i 的像素的亮度值, 边的集合 $E = \{(V_i, V_j) \mid \|P_i - P_j\| \leq d\}$ 。

按照以上原则构建了图像的图的存储结构之后, 必然可以得到图的一些连通分量, 这些连通分量对应着图像中不规则块。这里存在的一个问题, 创建图的时候, 当给定的颜色距离阈值过小时, 会造成图像过度分割, 也就是说对应图像的连通分量的个数会大增; 当给定的颜色距离阈值过大时, 又会造成分割不力的现象, 这时, 一个连通分量可能包含图像的多个语义块。针对这两种现象有不同的处理办法, 对第一种情况, 可以采用合并连通分量的办法, 使其成为一个完整的语义块; 对第二种情况, 则可以再进行再分割, 直至得到比较合理的图的划分时结束。本文的方法采用了第一种处理办法即小阈值分割, 然后进行不规则块的合并。对不规则块的合并, 主要是针对一个语义块被分成了几个连通分量的情况, 如: 运动员的背部和腹部由于光照和遮挡的影响, 在创建图的时候被划分成了两个不同的连通分量, 但其同属于运动员的躯体, 应该

当作一个语义块来进行处理。对于这种情况, 本文采用假设检验的方法来度量区域的相似性。

假设两个相邻的区域 R_1 和 R_2 分别包含有 n_1, n_2 个点, 有两种可能的假设:

H_0 : 两个区域属于同一物体。在这种情况下, 两个区域的颜色值都服从单一高斯分布 (μ_0, σ_0^2) 。

H_1 : 两个区域属于不同的物体。在这种情况下, 每一个区域的颜色值服从不同的高斯分布 (μ_1, σ_1^2) 和 (μ_2, σ_2^2) 。

$$\begin{aligned} P(a_1, a_2, \dots, a_{n_1+n_2} | H_0) &= \prod_{i=1}^{n_1+n_2} P(a_i, H_0) \\ &= \prod_{i=1}^{n_1+n_2} \frac{1}{\sqrt{2\pi\sigma_0}} e^{-\frac{(a_i - \mu_0)^2}{2\sigma_0^2}} = \frac{1}{(\sqrt{2\pi\sigma_0})^{n_1+n_2}} e^{-\frac{\sum_{i=1}^{n_1+n_2} (a_i - \mu_0)^2}{2\sigma_0^2}} \\ &= \frac{1}{(\sqrt{2\pi\sigma_0})^{n_1+n_2}} e^{-\frac{(n_1+n_2)\mu_0^2}{2\sigma_0^2}} \end{aligned} \quad (11)$$

$$\begin{aligned} P(a_1, a_2, \dots, a_{m_1}, a_{m_1+1}, \dots, a_{m_1+m_2} | H_1) \\ &= \frac{1}{(\sqrt{2\pi\sigma_1})^{m_1}} e^{-\frac{m_1\mu_1^2}{2\sigma_1^2}} \frac{1}{(\sqrt{2\pi\sigma_2})^{m_2}} e^{-\frac{m_2\mu_2^2}{2\sigma_2^2}} \end{aligned} \quad (12)$$

似然比

$$L = \frac{P(a_1, a_2, \dots, | H_1)}{P(a_1, a_2, \dots, | H_0)} = \frac{\sigma_0^{n_1+n_2}}{\sigma_1^{m_1} \cdot \sigma_2^{m_2}} \quad (13)$$

通过计算式(13)的似然比 L 来决定是否合并这两个区域。

2.4 不规则块匹配

经过前两步之后, 我们已经获得了前一帧的前景分割图像。并且由于图像是以图的形式存储的, 那么分割得到的前景图像可能包括几个连通分量。一般地, 大致可以分为头部, 裸露的皮肤以及着装的躯体三大部分。不规则块匹配就是要根据前一帧的分割结果在下一帧指定的 ROI 内寻找相匹配块。这其实也就是相似性比较的问题, 经典的相似性度量规则见文[13~17], 绝对差(AD)、平均绝对差(MAD)、平方差(SD)、平均平方差(MSD)、积相关(Prod)和归一化积相关(Nprod), 前四个为最小相似性度量, 而 Prod 和 Nprod 为最大相似性度量。本文借鉴了变量间的相似性度量准则^[18]——相似系数法来进行不规则块匹配的相似性度量。一般地, 将前一帧分割得到的前景区域的某一部分所对应的连通分量记为变量 X_α , 假设下一帧 ROI 中有一连通分量 X_β , 则 X_α, X_β 之间的相关系数定义为:

$$r_{\alpha\beta} = \frac{S_{\alpha\beta}}{\sqrt{S_{\alpha\alpha}S_{\beta\beta}}} = \frac{\sum_{i=1}^n (X_{\alpha_i} - \bar{X}_\alpha)(X_{\beta_i} - \bar{X}_\beta)}{\sqrt{\sum_{i=1}^n (X_{\alpha_i} - \bar{X}_\alpha)^2 \sum_{i=1}^n (X_{\beta_i} - \bar{X}_\beta)^2}} \quad (14)$$

其中, $(X_{1\alpha}, X_{2\alpha}, \dots, X_{n\alpha})$ 为 X_α 中所有像素的取值, \bar{X} 为 X_α 中所有像素的平均值, $r_{\alpha\beta}$ 的绝对值越接近 1, 说明块 X_α, X_β 的相似性越高。一般地, $r_{\alpha\beta}$ 应满足:

$$1) |r_{\alpha\beta}| \leq 1 \text{ 且 } r_{\alpha\alpha} = 1; 2) r_{\alpha\beta} = \pm 1 \Leftrightarrow X_\alpha = X_\beta (C \neq 0); 3) r_{\alpha\beta} = r_{\beta\alpha}.$$

由于事先已经将图像存储为图的形式, 不规则块匹配的问题就可以转换为比较子图的相似性, 采用深度优先遍历的方法(DFS)遍历块 X_α, X_β 中的每一个像素, 并按照式(14)计算 $r_{\alpha\beta}$ 的值, 在下一帧中找到一个 X_β , 使得其和 X_α 的 $r_{\alpha\beta}$ 值最大, 即为与 X_α 相匹配的块。如图(3)所示, 图 3 为上一帧分割得到的前景, 图 4 为经过区域合并得到的图像划分, 蓝色方框为 ROI。



图 3 前一帧分割得到的前景

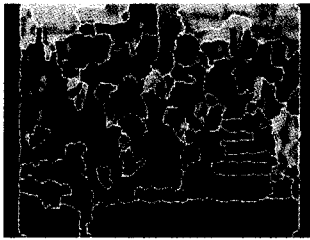


图4 当前帧图像 ROI

2.5 后处理

将分割得到的结果进行二值化后,前景区域有可能出现空洞并且有零散噪声的影响。图5为未进行后处理的图像与经过后处理的分割结果的对比。对于零散的杂点噪声,本文采用连通成分分析方法将其消除。具体操作的时候,通过计算前景中所有连通区域面积,将面积小于整个图像面积2%的连通区域消除。运用数学形态的膨胀运算可以很好地填充前景中的空洞区域。

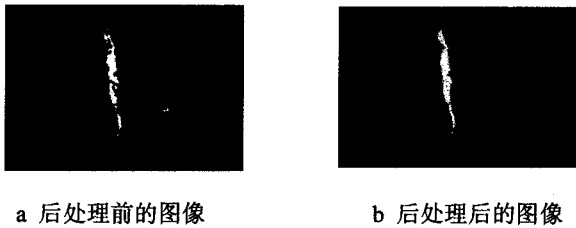


图5 后处理结果

3 实验结果

为了检验本文算法的有效性,本文对几种不同的跳水视频序列进行了测试,测试序列1(Diving1)的图像大小为 352×240 ,测试序列2(Diving2)的图像大小为 320×240 。算法的执行速度为每帧平均耗时约100ms,可以达到实时分割的目的。

图6为一段十米跳台跳水视频及其分割结果。图7为3米跳板跳水视频及分割结果。在这两个序列中,前景和背景都存在运动。其中,第一段视频运动比较缓慢,前后两帧的变化不会超过 ± 8 个像素,第二段视频比较短,变化快,前后两帧的变化大概在 ± 12 左右。

实验结果表明,本文所提出的方法能够有效地从动态场景视频序列中检测出肤色裸露、着装比较单一的运动前景,这种方法比较适合对体育运动中的视频,如:跳水、体操等视频的运动前景对象提取。

结论 本文针对肤色裸露、着装单一的这一类存在全局运动的视频,提出了一种基于局部图金字塔的不规则块匹配的视频分割方法。该方法通过构造图像ROI中的图的存储结构,可以将运动前景划分为几个连通分量,根据上一帧分割得到的前景块,通过不规则块匹配的方法找到当前帧中的前景图像块。该方法主要适用于体育运动中的跳水、体操视频的视频分割。我们下一步的工作将考虑对于任意的存在全局运动的视频,在只有第一帧需要人工交互的前提下,能实时得到比较精确的分割结果。

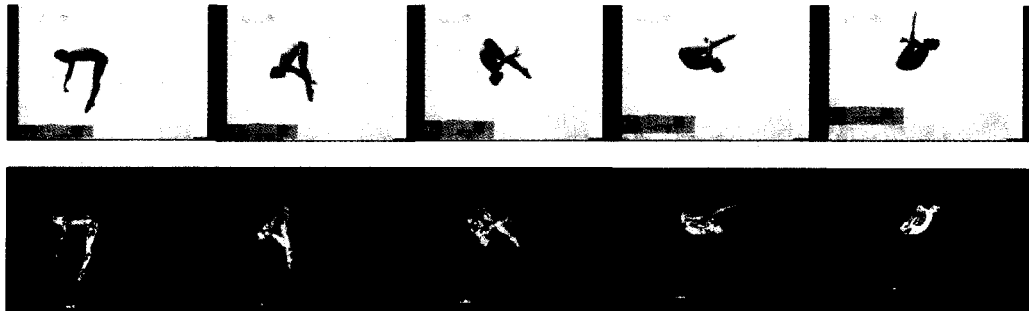


图6 十米跳台跳水视频分割(上图:原始视频,下图:分割结果)

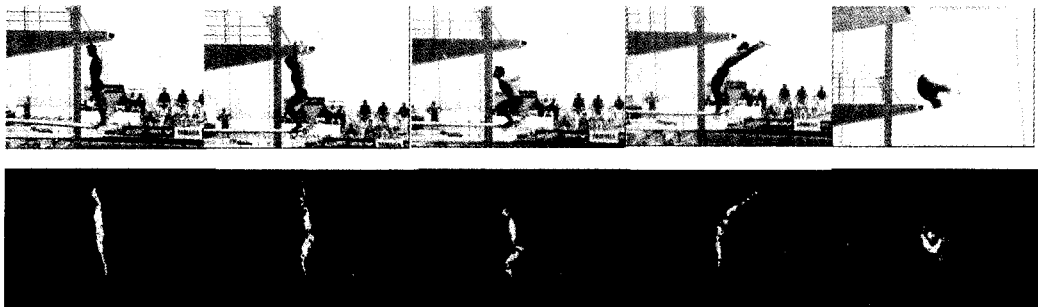


图7 3米跳板跳水视频分割(上图:原始视频,下图:分割结果)

参考文献

- 1 Farin D, de With P H N, E# elsborg W. Video-object segmentation using multi-sprite background subtraction. In: Proc. IEEE International Conference on Multimedia and Expo (ICME), 2004
- 2 吴思,林守勋,张勇东. 基于动态背景构造的视频运动对象自动分割. 计算机学报, 2005(8):1386~1392
- 3 Bergen J R, Burt P J, Hingorani R, Peleg S. A Three-Frame Algorithm for Estimating Two-Component Image Motion. IEEE Abstract. IEEE Top Reference. WWW Version. BibRef 9209. PAMI, 1992,14(9):886~896
- 4 Teixeira L F, Corte-Real L. Cascaded change detection for foreground segmentation. FEUP / INESC Porto. Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal
- 5 Yu T, Zhang C, Cohen M, et al. Monocular Video Foreground/

- Background Segmentation by Tracking Spatial-Color Gaussian Mixture Models. IEEE Workshop on Motion and Video Computing 2007, Austin, TA, Feb. 2007
- 6 Wang Y, Loe K, Tan T, et al. Spatiotemporal Video Segmentation Based on Graphical Models. IEEE Trans. On Image Processing, 2005, 14(7): 937~947
 - 7 Chien S, Huang Y, Hsieh B, et al. Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques. IEEE Trans. on Multimedia, 2004, 6(5): 732~748
 - 8 Liu Y, Zheng Y F. Video Object Segmentation and Tracking Using Learning Classification. IEEE Trans. on Circuits System and Video Technology, 2005, 15(7): 885~899
 - 9 Chen P C, Su J J, Tsai Y P. Coarse-To-Fine Video Object Segmentation By MAP Labeling of Watershed Regions, Bulletin of College of Engineering, National Taiwan University, 2004(9): 25~34
 - 10 Malik J, Shi J. Normalized cuts and image segmentation. In: Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, 1997
 - 11 Feleznszwalb P, Huttenlocher D. Image segmentation using local variation. In: Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, 1998. 98~104
 - 12 Cormen T H, Leiserson C E, Rivest R L, Stein C. Introduction to Algorithms, Second Edition. MIT Press and McGraw-Hill, ISBN 0-262-03293-7. Chapter 26; Maximum Flow, 2001. 643~700
 - 13 孙仲康, 沈振康. 数字图像处理及其应用[M]. 北京: 国防工业出版社, 1985
 - 14 Zitova B, Flusser J. Image Registration Methods; A Survey[J]. Image and Vision Computing, 2003, 21 (11): 977~1000
 - 15 Lisa G B. A Survey of Image Registration Techniques[J]. ACM Computing Surveys, 1992, 24 (4): 325~376
 - 16 曹菲, 杨小冈, 缪栋, 等. 景象匹配制导基准图选定准则研究[J]. 计算机应用研究, 2005, 22 (7): 137~139
 - 17 杨小冈, 曹菲, 缪栋, 等. 基于相似度比较的图像灰度匹配算法研究[J]. 系统工程与电子技术, 2005, 27(5): 918~921
 - 18 Cohen J, Cohen P, West S G, Aiken L S. Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, 2003

(上接第 230 页)

我们对多个视频序列(均为 320×240 的分辨率)采用本文提出的算法进行测试,包括室外场景和室内场景,每个场景都包含噪声运动区域。为了提高检测速度,采用灰度图进行计算。图 3 显示了其中的 4 个测试视频及其检测结果。图 3(a)显示某个场景中存在一个摆动的植物,经过背景模态检测后,植物的枝叶部分被划分成多模态;图 3(b)显示一个包含大量树木的室外场景,这是一种比较极端的情况,经过背景模态检测,运动的植物区域被分类出来;图 3(c)显示一个室内场景,只有电视机部分以及一些反光的边缘被划分为多模态背景区域;图 3(d)显示存在一条河流的场景,河流波动较大的区域被划分成多模态背景区域。

总结 与文[13]提出的背景分类方法相比,本文提出的方法在速度上有着明显的优势,平均在数分钟内就能完成计算,而文[13]提出的基于神经网络的方法需要数小时乃至数天的运算时间。由于在大部分监控应用中,所监控的场景都不是频繁变化的,所以本文的方法能够满足实际的需求。

对于背景更新问题,所提出的方法都是采用单一模型对所有像素进行更新,没有考虑到背景中像素的模态性差异。很多复杂的方法尽管能够描述多模态背景像素值的分布,但是计算复杂,难以应用于实际。我们认为:由于一般情况下,背景中的动态区域所占背景的比例较小,所以没有必要为所有像素采用同一个复杂的更新方法,所以事先对背景像素进行模态检测就可以为后续的背景更新操作提供方法选择上的依据,从而在不降低精确度的同时大幅提高背景更新速度。下一步的工作是优化该算法,使其运算速度进一步提高。

参 考 文 献

- 1 Toyama K, Krumm J, Brumitt B, et al. Wallflower: principles and practice of background maintenance. In: Proceedings of the IEEE International Conference on Computer Vision. Kerkyra Greece, 1999. 255~261
- 2 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, 1999. 23~25
- 3 Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction. In: Proceedings of the European Conference on Computer Vision. Dublin Ireland, 2000. 751~767
- 4 侯志强, 韩崇昭. 基于像素灰度归类的背景重构算法. 软件学报, 2005, 16(9): 1569~1576
- 5 Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809~830
- 6 Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using Kalman-filter. In: Proceedings of the International Conference on Recent Advances in Mechatronics. UNESCO Chair on Mechatronics, 1995. 193~199
- 7 Fukunaga K, Hostetler L D. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans on Information Theory, 1975, 21(1): 32~40
- 8 Cheng Y. Mean shift, mode seeking and clustering. IEEE Trans on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790~799
- 9 Comanicu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Trans Pattern Analysis and Machine Intelligence, 2002, 24(5): 603~619
- 10 Meer P, Georgescu B. Edge detection with embedded confidence. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(12): 1351~1365
- 11 Christoudias C, Georgescu B, Meer P. Synergism in low level vision. In: Proceedings of the 16th International Conference of Pattern Recognition. Quebec City, Canada, 2001. 150~155
- 12 Parzen E. On the estimation of a probability density function and mode. Ann Math Stat, 1962, 33: 1065~1076
- 13 Gil-Jim'enez P, Maldonado-Basc'on S, Gil-Pita R, et al. Background Pixel Classification for Motion Detection in Video Image Sequences. Lecture Notes in Computer Science, 2003, 2686: 718~725
- 14 Wren C, Azarbayejani A, Darrell T, et al. Pfister: Real-time Tracking of the Human Body. IEEE Trans on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780~785
- 15 Scott D W. Multivariate Density Estimation. Wiley-Interscience, 1992
- 16 Turlach B A. Bandwidth selection in kernel density estimation: A review. Discussion Paper. 9317, Institut de Statistique, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, 1993