

一种用于跳水运动视频的全局运动估计方法

邹北骥^{1,2}, 韩立芹¹, 彭小宁²

(1. 中南大学信息科学与工程学院, 湖南长沙 410083; 2. 怀化学院计算机系, 湖南怀化 418008)

摘要: 提出一种适用于跳水运动视频的全局运动估计方法。跳水运动中的运动员着装简单、肤色遮挡较少, 通过测试阈值建立肤色模型, 快速提取前景和背景中的人体区域, 避免其对全局运动参数估算的影响; 利用帧差法去除相邻图像帧内相对静止的光流场, 根据参照宏块运动方向进一步去除影响全局运动估计的宏块, 减少宏块匹配时间。实验结果表明, 算法不但简化了全局运动参数估算的复杂性, 而且提高了的计算精度。

关键词: 全局运动估计; 肤色模型; 运动向量; YCbCr 颜色空间; 参照宏块

中图分类号: TP301. 6 **文献标识码:** A **文章编号:** 0372-2112 (2008) 12-2412-06

A Global Motion Estimation Method for Diving Video Sequences

ZOU Bei-ji^{1,2}, HAN Li-qin¹, PENG Xiao-ning²

(1. School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China;

2. Department of Computer Science and Technology, Huaihua University, Huaihua, Hunan 418008, China)

Abstract: This paper presents a global motion estimation method for diving video. Human body color model is constructed base on the athletes' simple dress in diving video and the area of human body within the foreground and background is picked up so as to improve the efficiency of global motion estimation. Meanwhile, the adjacent frame difference and the direction of the reference module are used to remove the modules which not used in the global motion estimation, so module matching time is reduced accordingly. Experimental results show that the presented method can not merely reduce the computational complexity of the global motion estimation's, but also improve its accuracy.

Key words: global motion estimation; body color model; motion vector; YCbCr color space; reference module

1 引言

跳水运动视频中, 存在两种运动: 一是由摄像机的变化(如摇摄、平移等)而引起的整个图像的变化, 即全局运动; 另一是由运动员运动而引起的局部变化, 即局部运动。获得准确的摄像机的全局运动参数是进行视频分析的前提和基础。

对于全局运动估计, 有不同的参数模型用来表示摄像机的镜头变化: 用于 MPEG-4 编码的八参数透视模型^[4]囊括了镜头的全部变化; 六参数的仿射模型^[5]表示镜头的摇摄和倾斜、在 x 轴和 y 轴的旋转以及变焦; 四参数的模型^[6,7]表示镜头的摇摄、变焦以及镜头在 x 轴和 y 轴的旋转。参数模型使用的参数越多, 估计全局运动越准确, 但计算也越复杂。同时有不同计算相邻两帧图像的光流场变化方式: 基于像素灰度值对应的像素法^[1]、基于图像中一小块区域的宏块法^[2]、基于图像中典型特征的特征法^[3]等。进行全局运动估计时, 场景中运动对象

引入了全局运动参数估计误差。常用的方法是在进行参数估计之前对图像进行判别, 挑选部分图像用于参数估计^[2], 也可以在参数估计过程中对图像进行判别^[7]。

跳水运动视频中, 摄像机的旋转变化比摇摄、变焦变化出现的概率小, 本文采用了文献[8]的基于宏块的三参数模型法来计算跳水运动视频的全局运动估计参数。由于视频中人体以及跳板等运动对全局运动估计的影响, 提出了适用于跳水运动视频的全局运动估计方法。主要贡献在于: (1) 根据肤色模型, 提取出前景的人体区域以及背景中的杂乱运动人体, 避免了前景运动对全局运动参数的估计误差; (2) 由于全局运动参数计算只涉及图像帧内变化的光流场, 利用帧差法去除图像帧内静止的光流场, 避免了无效的宏块匹配时间; (3) 提出了参照宏块的思想, 自动选定背景中一个宏块, 计算该宏块的运动方向, 参照该宏块的运动方向, 判断其它宏块是否需要参与计算。实验表明该算法不仅提高了运动估计的精度, 而且简化了计算的复杂度。

收稿日期: 2007-09-11; 修回日期: 2008-04-20

基金项目: 国家自然科学基金(No. 60673093); 国家自然科学基金重大研究计划(No. 90715043); 湖南自然科学基金(No. 06JJ2065); 长江学者和创新团队发展计划(No. IRT0661)

2 总体思路

本文采用宏块法计算摄像机全局运动参数, 基本流程如图 1 所示。 I_{t-1} 和 I_t 分别表示视频中的前一帧和当前帧, 首先取得当前帧 I_t 中的预匹配宏块, 在前一帧 I_{t-1} 中搜索, 得到最佳匹配宏块, 求得该宏块的运动向量, 根据三参数模型^[8]的公式计算全局运动参数, 从而确定 I_t 中各像素点相对于 I_{t-1} 的全局运动坐标。

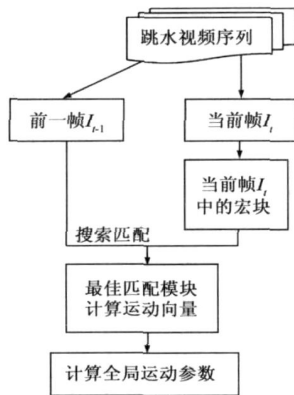


图1 全局运动估计总体流程

本文的全局运动参数计算由三个部分组成:

(1) 取得 I_t 中的预匹配宏块. 图 2(a), 图 2(b) 中所示的跳水运动视频中前景和背景中人体的运动都会导致全局运动计算误差. 本文采用肤色模型提取出视频前景和背景中的人体, 避免人体运动导致的误差. 另外静止的图像区域对全局运动参数的估计是无用的, 但却耗费宏块匹配时间, 本文利用帧差法去除静止的图像区域, 从而节省宏块匹配的时间。

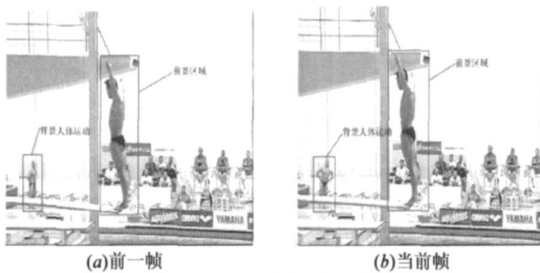


图2 跳水运动视频中杂乱的背景

(2) 匹配宏块. 由实验知, 跳水运动视频连续间摄像机位移不是特别大, 本文只在前一帧中包含该宏块左上角坐标的一个小范围内进行全搜索匹配以求得最佳匹配宏块。

(3) 全局运动参数的计算. 由当前帧的有效宏块数组 $block[n]$ 和其在前一帧的最佳匹配宏块数组 $match[n]$ 便可计算出每个宏块的运动向量. 如图 3(a)、(b) 所示, 跳水运动视频中, 跳板和前景运动趋势一致, 可认

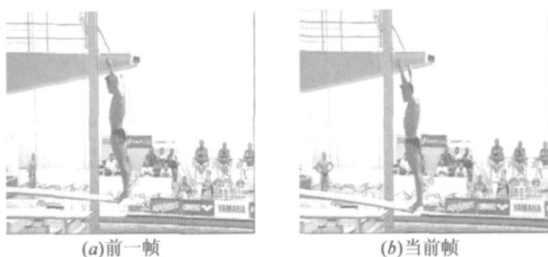


图3 跳板的运动

为跳板是前景. 根据相对运动即前景和背景的运动方向相反的特点, 本文利用了参照宏块的思想. 通过选定背景中的一个宏块作为参照宏块, 根据参照宏块的运动方向, 去除跳板区域。

3 全局运动参数计算过程

3.1 人体肤色模型

前景和背景中人体的运动将导致全局运动参数估计的误差, 因此需要首先剔除人体区域. 跳水运动视频中, 运动员肤色遮挡较少, 可通过肤色模型提取出人体. 本文采用 YCbCr 颜色空间^[9], 该颜色空间的优点是受亮度变化的影响较小, 且二维独立分布, 能较好地限制肤色分布区域. 在 YCbCr 颜色空间中, 本文针对 1000 个视频中人体肤色的 YCbCr 颜色值进行统计, 主要集中在图 4 所示的区域中。

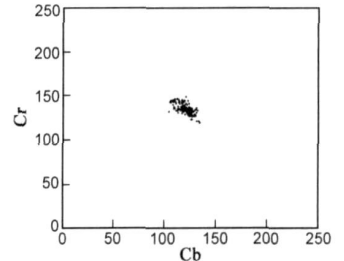


图4 人体肤色分布

限制 Cb、Cr 值则可确定人体区域. 实验验证, 采用 $Y \in [130, 165]$, $Cb \in [108, 129]$, $Cr \in [127, 146]$, 有效地提出了图像中的人体区域。

分别对 I_t 和 I_{t-1} 根据人体肤色模型进行预处理, 提取出人体并进行闭合操作以得到连通图像。

3.2 宏块数组

I_t 和 I_{t-1} 经过预处理之后, 由公式^[8]知, 全局运动参数估计只用到了光流相对变化的图像区域, 因此根据帧差法去除静止的图像区域。

本文通过设立阈值的方法来判断宏块是否有足够的梯度信息来估计全局运动参数, 从而避免光照等因素的影响. 如果宏块的梯度信息小于该阈值, 则为“不可靠宏块”, 不予考虑. 图像梯度信息本文采用了 sobel 算子, 每个宏块的大小为 $16 * 16$. 设 $P_{i,j}$ 为某一宏块中坐标为 (i, j) 的像素, $G(P_{i,j})$ 为该像素的梯度值^[8], 则该宏块的梯度值 G 为:

$$G = \sum_{i=1}^{16} \sum_{j=1}^{16} [G(P_{i,j}) / (16 * 16)]$$

宏块尺寸减小时, 越倾向于一个像素, 宏块的数量也相应增加, 势必会增加计算复杂度, 且对噪声越加敏感; 相反宏块尺寸增大时, 宏块数量会减少, 则计算结果会依赖这些较少宏块的值. 宏块大小(本文采用 $16 * 16$)固定时, 设立的阈值越大, 可靠宏块越少, 计算结果越依赖于可靠宏块的值; 设立的阈值越小, 可靠宏块越多, 将增加计算的复杂性. 由文献^[8]知, 可靠宏块的数目应保证在总宏块数量的 30% 以上. 本文是在帧差图

上进行宏块预判, 宏块数目可略小于 30%. 图 5(a) 显示了阈值大小与宏块数量的关系, 图 5(b) 显示了阈值大小与计算时间的关系. 根据实验统计, 本文采用的阈值为 25, 可满足计算的时间和精度要求.

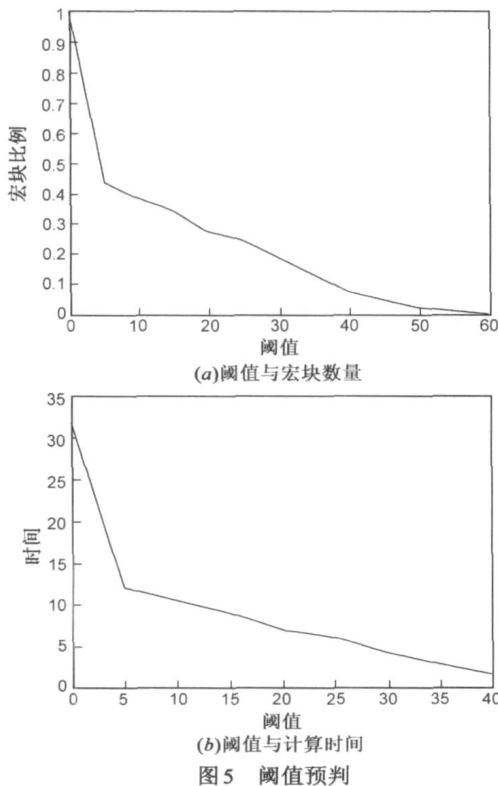


图5 阈值预判

利用 sobel 算子求得 I_t 与 I_{t-1} 的帧差图的梯度图, 分割成大小为 16×16 的宏块, 分别计算每个宏块的梯度值 G , 当 $G > 25$ 则为可靠宏块, 把宏块左上角的坐标存入宏块数组 $block[n]$, 反之则为不可靠宏块, 不操作.

如图 3(a), (b) 所示, 跳板视为前景, 可由前景和背景运动方向相反的特点确定宏块是背景或前景, 如图 6.

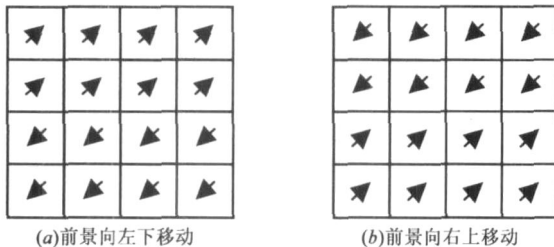


图6 图像光流的运动

本文利用了参照宏块的思想. 首先在宏块数组中选择一个不包括跳板的宏块 $block$, 利用匹配原则得到其在前一帧中匹配宏块 $match$, 求得该宏块的运动向量 V

$$V = block - match$$

然后对宏块数组 $block[n]$ 中的每一个宏块分别在前一帧根据匹配原则进行搜索匹配, 求得最佳匹配宏块, 计算其运动向量 v . 如果 v 与 V 速度方向相同为背

景宏块, 记录该宏块及其匹配宏块, 反之为前景宏块, 不参与计算.

本文采用了基于梯度信息和灰度信息的块匹配原则. 在前一帧 I_{t-1} , $g(i, t-1) = \{g(x, y, t-1)\}$ 表示第 i 个宏块内各个像素的梯度, $d(i, t-1) = \{d(x, y, t-1)\}$ 表示第 i 个宏块内各个像素的灰度. 在当前帧 I_t , 待匹配的宏块 $block$, $g(t) = \{g(bx, by, t)\}$ 表示宏块 $block$ 内各个像素的梯度, $d(t) = \{d(bx, by, t)\}$ 表示宏块 $block$ 内各个像素的灰度, 则匹配原则:

$$\varepsilon = |g(i, t-1) - g(t)| + |d(i, t-1) - d(t)|$$

搜索前一帧 I_{t-1} 的某一块区域, ε 最小的区域则为 $block$ 的最佳匹配宏块.

3.3 全局运动参数计算

跳水运动视频中, 由于摄像机的旋转与摇摄、变焦相比, 出现的概率非常小, 因此本文采用了文献[8]中的三参数全局运动估计模型.

以图像的中点作为原点, 假设某像素相对于图像中点的坐标为 (x, y) , 运动向量为 (v_x, v_y) , 则三参数运动模型可表示为:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = a \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_x \\ a_y \end{bmatrix} \quad (1)$$

其中 $a = z_{xy}$

$$a_x = f_x(p_x, z_{xy})$$

$$a_y = f_y(p_y, z_{xy}) \quad (2)$$

z_{xy} 是摄像机的变焦(缩放)系数, 而 (p_x, p_y) 则是摄像机摇摄时在 x 和 y 轴的平移分量.

有了以上的模型, 可通过 3 个模型参数来描述全局运动. 本文按照两步估计全局运动参数: 第一步, 将图像分割成若干个矩形宏块, 并利用块匹配算法计算光流场; 第二步, 根据得到的光流场, 使用最小二乘法计算 3 个全局运动参数.

假设当前帧被分割成 N 个宏块, 且每个宏块的运动向量为其中央像素的运动向量, 对于第 k ($k = 0, 1, \dots, N-1$) 个宏块, 用 (v_x^k, v_y^k) 表示经过块匹配算法得到的运动向量, (s_x^k, s_y^k) 表示其中央像素的坐标(相对于图像中点), 则三参数全局运动模型可写为

$$\begin{bmatrix} v_x^k \\ v_y^k \end{bmatrix} = a \begin{bmatrix} s_x^k \\ s_y^k \end{bmatrix} + \begin{bmatrix} a_x \\ a_y \end{bmatrix} \quad (3)$$

下面, 按照如下过程估计 3 个全局运动参数.

首先根据最小二乘原则, 定义指标函数

$$f(a, a_x, a_y) = \sum_{k=0}^{N-1} \left[(v_x^k - a s_x^k - a_x)^2 + (v_y^k - a s_y^k - a_y)^2 \right] \quad (4)$$

则 (a, a_x, a_y) 应使指标函数 $f(a, a_x, a_y)$ 达到最优:

$$(a, a_x, a_y) = \arg \min_{a, a_x, a_y} \sum_{k=0}^{N-1} \left[(v_x^k - a s_x^k - a_x)^2 + (v_y^k - a s_y^k - a_y)^2 \right] \quad (5)$$

解方程可以得到最优的 3 个全局运动参数 (a , a_x , a_y).

在本文算法中, 首先对当前帧和前一帧进行预处理, 并去除了静止的图像区域, 而在跳水运动视频中, 摄像机时而静止时而运动, 因此必然存在某一时刻, 由于光照等因素的影响引起一定量的宏块, 但是宏块的数量较少, 若根据公式计算全局运动参数, 则计算结果势必依赖这些宏块而不准确. 本文中为了避免这个误差, 采用了设定阈值的方法, 即当宏块数量小于这个阈值时, 则认为摄像机静止. 图 7 显示了阈值与计算结果的关系, 根据实验统计本文的阈值取为 10. 即当宏块数量小于 10 时, 摄像机静止, 有 $(a, a_x, a_y) = (0, 0, 0)$.

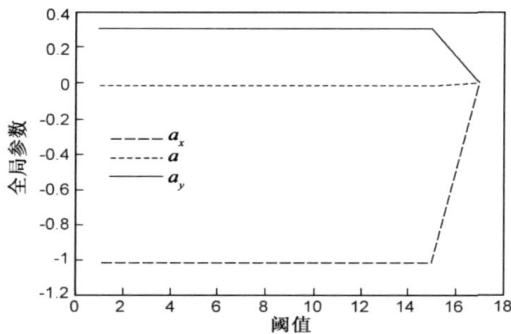


图 7 阈值与计算结果

4 实验结果及讨论

我们利用 matlab7.04 在 Pentium(R) 4 CPU 3.00 GHz, 512MB 的计算机上实现了本文的算法. 为了更好的说明本文算法的有效性, 分别对计算过程以及计算结果分析讨论如下:

4.1 摄像机静止的情况

摄像机静止时, 全局运动参数 $(a, a_x, a_y) = (0, 0, 0)$. 图 8~ 13 显示了当全局运动参数 (a, a_x, a_y) 为 $(0, 0, 0)$ 时的计算过程.

图 8 显示了输入视频中当全局运动参数 (a, a_x, a_y) 为 $(0, 0, 0)$ 的两帧:

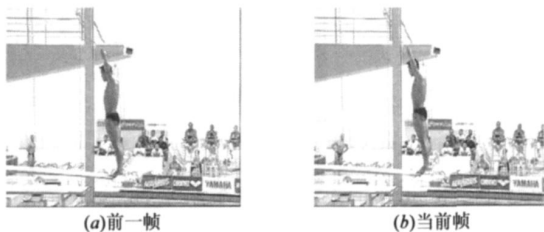


图 8 原图

图 9 是当前帧 I_t 和前一帧 I_{t-1} 的预处理图像. 由图像可知人体肤色模型较好地提取出了前景和背景中的人体区域.

图 10 是当前帧 I_t 和前一帧 I_{t-1} 的帧差图. 由图像可看出帧差图去除了图像中静止的区域. 图 11 是帧差

图的 sobel 图, 并由梯度信息确定的有效宏块, 如图 12 所示.

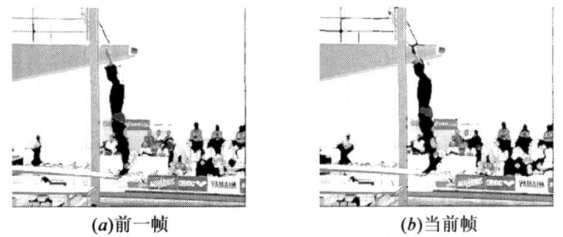


图 9 预处理后灰度图

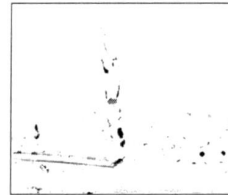


图 10 帧差图



图 11 帧差图的 sobel 图

图 12 显示了摄像机静止时, 有效宏块的个数较少, 则宏块匹配的时间也较短. 图 13 显示了宏块数组及其匹配数组. 其中红色的圆圈表示参与计算的匹配宏块, 绿色圆圈表示与参照宏块速度方向不一致而没有参与计算的匹配宏块.

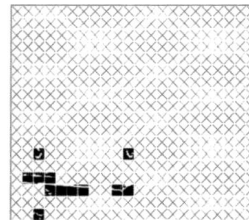


图 12 有效宏块图

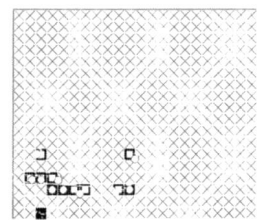


图 13 有效宏块匹配图

4.2 摄像机运动的情况

摄像机运动时, 全局运动参数 (a, a_x, a_y) 是变化的, 图 14~ 19 显示了全局运动参数变化时的计算过程, $(a, a_x, a_y) = (0.00, -0.65, 2.81)$

图 14 显示了输入视频中摄像机运动时的两帧:

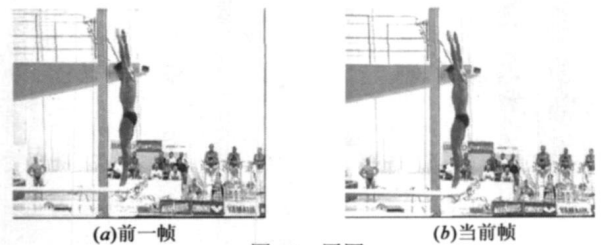


图 14 原图

图 15 是当前帧 I_t 和前一帧 I_{t-1} 的预处理图像. 由图像可知人体肤色模型较好地提取出了前景和背景中的人体区域.

图 16 是当前帧 I_t 和前一帧 I_{t-1} 的帧差图. 由图像可看出帧差图去除了图像中静止的区域. 图 17 是帧差图的 sobel 图像, 并由梯度信息确定的有效宏块, 如图 18

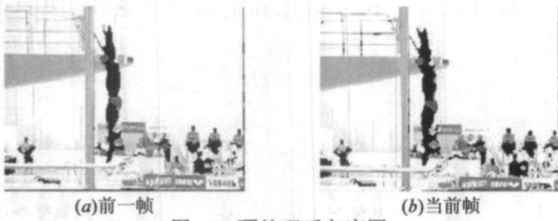


图 15 预处理后灰度图

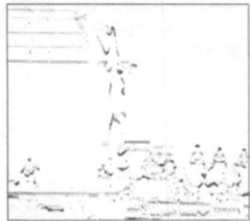


图 16 帧差图

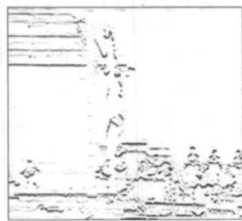


图 17 帧差图的sobel图

所示。

图 18 显示了在摄像机运动时,有效宏块的个数相比摄像机静止时较多。图 19 中显示了宏块数组及其匹配数组。

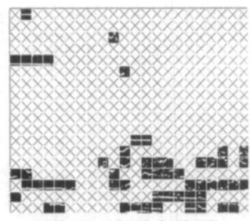


图 18 有效宏块图

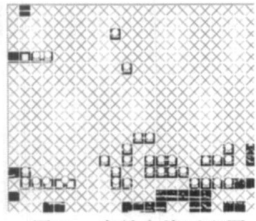


图 19 有效宏块匹配图

4.3 讨论

本文分析讨论了摄像机静止和运动时的全局运动参数的计算过程及其结果。表 1 显示了摄像机各情况的各步的计算时间。表 2 显示了摄像机各情况的计算结果,文献[8]方法的计算时间和结果如表 3 和表 4 所示。

表 1 摄像机各情况的各步运行时间

时间/(ms)	预处理	宏块预判	宏块匹配	全局参数计算	总时间
静止	44.8	8.4	1446.0	...	1449.2
运动	46.1	8.7	4276.6	0.17	4331.57

表 2 摄像机各情况的计算结果

参数	a	a_x	a_y
静止	0.00	0.00	0.00
运动	0.00	-0.65	2.81

前景区域不仅影响图像匹配时间,如表 3,而且也导致全局运动参数估计的误差,如表 4。在本文中,去除了前景区域,并去除静止的图像区域,节省了宏块匹配的时间,同时为了避免光照等因素引起的少量区域变

化,设定了一个阈值。摄像机静止时宏块的个数小于阈值,直接令 (a, a_x, a_y) 为 $(0, 0, 0)$, 由此避免了因少量宏块的计算误差而引起的错误计算,提高了计算的精度。摄像机运动时宏块的个数大于阈值,则利用最小二乘法计算全局运动参数,由于去除了前景的影响,提高了计算的精度。如表 2 和表 4 所示。

表 3 文献[8]方法中摄像机各情况的各步运行时间

时间/(ms)	预处理	宏块预判	宏块匹配	全局参数计算	总时间
静止	...	9.0	7490.3	0.18	7499.48
运动	...	9.7	6774.9	0.20	6784.8

表 4 文献[8]方法中摄像机各情况的计算结果

参数	a	a_x	a_y
静止	0.00	-0.38	-0.84
运动	0.00	-0.95	2.02

从以上的分析和计算结果来看,得到的全局运动参数的数据以及计算的时间可以满足跳水运动视频的运动分析的要求。

5 结论

本文提出了基于肤色模型的适用于跳水运动视频的全局运动估计方法。前景的运动可引起摄像机全局运动参数估计的误差,在跳水运动视频中,运动员的肤色遮挡比较少,因此可以利用肤色模型提取出运动人体,避免了人体运动对全局运动参数估计的影响,从而提高运动估计的精度。另外在本文中根据相对论的中前景和背景的运动方向相反的思想,利用了运动向量的方向去除了前景区域,提高了计算的精度。另外根据全局运动参数估计的计算公式,可知图像中静止的区域对计算结果毫无用处,但却耗费大量的宏块匹配时间,因此本文首先利用相邻帧差法去除相对静止的图像区域,然后再进行全局运动参数估计计算。这样大大减少了宏块匹配的时间,大大简化了计算的复杂性。实验结果表明该方法是有效的。

参考文献:

[1] Ko S, Lee S, Jeon S, Kang E. Fast digital image stabilizer based on gray coded bit plane matching [J]. IEEE Transactions on Consumer Electronics, 1999, 45(3): 598-603.

[2] Sowar G, Murshed M, Doooley L. Fast global motion estimation using iterative least square technique [A]. Proceedings of the 4th International Conference on Information, Communications & Signal, Proceedings of the 4th IEEE Pacific Rim Conference on

- Multimedia[C]. Singapore, 2003. 282– 286.
- [3] Paik J K, Park Y C. An edge detection approach to digital image stabilization based on tri state adaptive linear neurons[J]. IEEE Transactions on Consumer Electronics, 1991, 37(3) : 521– 530.
- [4] M Irani, P Anandan, S Hsu. Mosaic based representations of video sequences and their applications[A]. International Conference on Computer Vision[C]. Boston, USA 1995. 605– 611.
- [5] F Moscheni, F Dufaux, M Kunt. A new two stage global/local motion estimation based on a back ground/ foreground segmentation[A]. IEEE Proceedings of International Conference on Acoustics, Speech and Signal Processing[C]. 1995. 2261– 2264.
- [6] S Wu, J Kittler. A differential method for simultaneously estimation of rotation, change of scale and translation[A]. Signal Processing: Image Communications[C]. IEEE Xplore, 1990. 69– 80.
- [7] G B Rath, A Makur. Iterative least squares and compression based estimations for a four parameter linear global motion compensation[J]. IEEE Transactions on Circuits & Systems for Video Technology, 1999, 9(7) : 1075– 1099.
- [8] 王嘉,王海峰,刘青山,卢汉清. 基于三参数模型的快速全局运动估计[J]. 计算机学报, 2006, 6(29) , 920– 927. Wang Jia, Wang Haifeng, Liu Qingshan, Lu Hanqing. Fast global motion estimation based on 3 parameter global motion model [J]. Chinese Journal of Computers, 2006, 6(29) , 920– 927. (in Chinese)
- [9] 梁路宏,艾海舟,徐光 ,张钊. 人脸检测研究综述[J]. 计算机学报, 2002, 5(25) , 449– 458. Liang Luhong, Ai Haizhou, Xu Guangyou, Zhang Bo. A survey of human face detection[J]. Chinese Journal of Computers, 2002, 5(25) , 449– 458. (in Chinese)

作者简介:



邹北骥 男, 1961 年生, 博士, 教授, 博士生导师, 主要研究领域: 计算机图形学、计算机视觉与数字图像处理等。

E mail: bjzou@vip. 163. com



韩立芹 女, 1982 年生, 硕士研究生, 主要研究领域: 计算机视觉与数字图像处理。

E mail: herry978@126. com