

一种面向快速全局运动估计的渐进精细阈值方法

郑嘉利^{1,2} 覃团发¹

(广西大学计算机与电子信息学院 南宁 530004)¹ (中国科学院计算技术研究所 北京 100190)²

摘要 快速全局运动估计的关键在于全局运动区域与局部运动区域的分割,其难点在于阈值的设定。提出一种面向快速全局运动的渐进精细阈值方法,该方法分为两步:第一步,用一个将统计特性与均值阈值相结合的亮度残差阈值模型来大致划分局部运动区域与全局运动区域,从而得到全局运动估计区域的近似集合;第二步,使用一种运动矢量残差分级阈值技术在能量残差函数最小化迭代过程中逐步细化全局运动像素点集合,最后分离出完整的全局运动区域,从而实现快速的运动估计。

关键词 全局运动估计, 阈值方法, 亮度残差, 运动矢量残差

中图分类号 TP391.41 **文献标识码** A

Progressive Fine Granular Threshold Method Oriented to Fast Global Motion Estimation

ZHENG Jia-li^{1,2} QI Tuan-fa¹

(School of Computer and Electronic Information, Guangxi University, Nanning 530004, China)¹

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)²

Abstract The method of the fast global motion estimation focuses on separating global motion region and local motion region. The key point of this method is how to set the appropriate threshold. This paper proposed a progressive fine granular threshold technique to solve this problem. The technique is divided into following two steps: firstly, distinguishing coarsely the global motion region and local motion region by using a luminance error threshold model integrating statistics method and average threshold method; secondly, during the iteration for minimizing error, refining the set of the global motion pixels by using a hierarchical threshold method based on motion vector error. By this means the whole global motion region can be obtained and the fast global motion estimation can be achieved.

Keywords Global motion estimation, Threshold method, Luminance error, Motion vector error

1 全局运动估计背景

在视频图像中,运动通常由摄像机运动和场景中的物体运动产生,由摄像机运动产生的运动将影响整个图像,称之为全局运动。物体的运动称为局部运动。如果视频图像可以用一个全局运动建模^[1],则用于视频编码时就可以节省大量的运动信息。全局运动估计^[2](Global Motion Estimation, GME)就是根据一定的镜头运动模型,利用视频序列中背景的运动信息得到该模型的具体参数。全局运动估计在视频压缩与编码、动态场景分析、视频内容分析^[3]和运动目标跟踪技术^[4]等数字视频处理方面有着广泛的应用。

但是,以往的研究表明,全局运动估计存在着计算量大的问题^[5]。全局运动估计关键是对全局运动模型参数的估计。理论上来说,对于一个确定的全局运动参数模型,如仿射六参数运动模型,只需3个背景像素点的运动矢量即可求得6个运动参数的值。但一般的运动矢量方法只能得到精度有限的运动矢量,如果仅用很少几个像素的运动矢量来估计全局运动模型参数,得到的结果精度非常低,因此需要更多像素的运动矢量参与全局运动模型参数估计来提高准确性。这样,参数估计就成了一个解矛盾方程组的问题,可用最小二乘法如

高斯牛顿迭代法^[6]来求解。通过数次迭代计算解方程组,直到最后全局运动模型参数值收敛于一个较为稳定的值。但这一迭代过程是全局运动估计中计算复杂度最高的部分,尤其在需处理的全局运动模型较为复杂、迭代次数较多的情况下。

再者,大多数的视频场景都包含有全局运动和局部运动,如果让所有像素的运动矢量都参与全局运动估计,则作局部运动的像素点以及运动矢量估计误差较大的像素点会对全局运动估计产生很大的干扰^[7],使得运动模型参数估计要被迭代很多次才能收敛于一个稳定值。这不但降低了全局运动估计的精度,而且使得运算复杂度大大升高。

局部运动区域里的像素点称为外点,全局运动像素点称为内点。我们要研究的快速全局运动估计方法的重点,就是把局部运动区域从整幅图像中有效分割出来,使得全局运动区域包含尽可能多的内点和尽可能少的外点。区分外点和内点的关键和难点在于阈值 T 的确定。在有些文献^[8]中, T 设置为一个固定的值(称之为固定阈值法),但是往往很难找到一个合适的固定阈值 T 胜任于整个迭代过程:如果 T 太小,迭代过程很可能一开始就收敛于需要局部运动区域的局部,尤其是局部运动区域占整个场景的比例较大的时候;而如果 T 太大则迭代结束的时候,全局运动像素点将有一部分被错

到稿日期:2010-07-21 返修日期:2010-12-23 本文受广西大学科研基金项目(XB2091010)资助。

郑嘉利(1979—),男,博士,副教授,主要研究方向为多媒体通信、视频编码;覃团发(1966—),男,博士,教授,主要研究方向为视频编码、视频检索。

判为局部运动点,从而大大影响参数估计的准确性。文献[9]中提到用百分比阈值 $r\%$ 代替阈值 T (称之为固定百分比阈值法),即令外点集合为残差值较大的 $r\%$ 像素的集合。但是由于不同视频中全局运动区域占整个场景的比例大小都不相同,因此这种方法要么只能得到部分的全局运动区域,要么在内点集合中会包含很多外点。文献[10]中提出了一种改进的方法:令阈值 T 为所有像素的残差的均值(称之为残差均值阈值法),但是残差均值法并不一定能很好地刻画全局运动区域和局部运动区域的差异,实验结果也表明在有些情况下这种方法可能使内点集合收敛于局部运动区域。

前面提到的这些内区域确定方法很难兼顾参数估计的准确性和鲁棒性,而且针对不同的视频场景往往需要事先选择一个理想的阈值,适应性较差。本文提出一种快速分割全局运动区域和局部运动区域的方法——渐进精细阈值法来克服这些缺陷。

2 亮度残差阈值法

在全局运动估计的初始阶段,我们提出一种基于亮度残差^[11]的阈值法来快速地大致区分全局运动区域和局部运动区域。该方法具有实现简单、算法复杂度低的特点,适合前景运动物体所占图像比例较小的情况。亮度残差均值阈值法的原理是设定阈值 C_g 为所有像素运动补偿后的亮度与相邻帧相同位置像素点亮度之间残差的均值,令亮度残差低于该阈值的像素点归为全局运动像素点集合,其余的是局部运动像素点集合。

2.1 亮度残差法

在运用递归最小二乘法估计全局运动时,初始化输入值是图像中各宏块经过块匹配运动估计后得到的运动矢量。处于局部运动区域的宏块的运动矢量不属于全局运动内点的集合,而是对全局运动参数估计造成干扰的外点,其将会消耗大量的运算时间。因此,如果在使用块匹配算法进行运动估计时就能够对宏块运动估计的属性进行预判,提前发现可能会干扰全局运动估计的宏块并予以剔除,必然会大大降低全局运动估计的计算量。

对视频图像的分析可以发现,全局运动区域的纹理较为丰富;而局部运动一般位于纹理灰度一致或灰度极其平滑的区域。这是因为局部运动区域所包含的运动信息比较少,相对来说,原始帧与参考帧对应块之间的亮度残差也很少。因此,本文通过宏块运动补偿后的亮度信息来对每个宏块进行预分析,以判断该宏块是否属于局部运动宏块,如果是,则把该宏块在进行全局运动估计之前从全局运动宏块集合中剔除出去,以节省全局运动估计的计算开销,提高全局运动估计的精度。

算法的主要思路是:首先用 sobel 算子^[12]分别计算当前块内所有像素的梯度均值 $S(i, j)$,其中, $S(i, j) = \sum_{i=1}^m \sum_{j=1}^n S_{i,j} / m \times n$,以及在参考帧中的对应块内所有像素的梯度均值 $S'(i, j)$,其中, $S'(i, j) = \sum_{i=1}^m \sum_{j=1}^n S'_{i,j} / m \times n$ 。比较 $S(i, j)$ 和 $S'(i, j)$ 残差绝对值的大小,也即 $|S(i, j) - S'(i, j)|$,如式(1)所示。

$$G_B = \sum_{i=1}^m \sum_{j=1}^n |S_{i,j} - S'_{i,j}| / m \times n \quad (1)$$

式中, $m \times n$ 是块大小。比较 G_B 与梯度阈值 C_g 的大小,把梯度均值残差小于阈值 C_g 的块从全局运动估计中剔除出去。

2.2 阈值 C_g 的推导

考虑到视频图像中纹理分布的不一致性以及前景运动物体所占比例的多样性,本文提出一种基于统计特性的亮度残差阈值法,其通过对主流的各种分辨率测试视频的实验验证,确定阈值因子。阈值 C_g 根据式(2)来确定:

$$C_g = k \cdot \frac{1}{N} \sum_{i=1}^N G_i \quad (2)$$

式中, N 是图像被划分成互不重叠的块的个数, G_i 是图像中第 i 个块的亮度残差(根据式(1)计算), $\frac{1}{N} \sum_{i=1}^N G_i$ 代表当前帧的亮度残差均值。如果当前块的亮度残差 G_B 小于 C_g ,则当前块被判定为局部运动区域块,从全局运动宏块集合中剔除出去。

该算法分为以下 3 步:

(1)把整幅图像分成互不重叠的块,块大小是 16×16 ,用块匹配算法对每一个块做运动估计,得到每一个块的运动矢量 $(\Delta x, \Delta y)$ 。理论上来说,块大小越小,运动估计结果越精确,一个极端的情况是当块大小为一个像素点时,运动估计结果最准确。但这样一来,运动估计的复杂度将大大升高,需要编码传输的运动矢量比特数也更多。本算法采用 16×16 的块,并兼顾考虑计算复杂度与运动估计精确性。同时,对于块匹配运动估计,我们采用的是三步搜索快速运动估计,以进一步降低运动估计的复杂度。通过块匹配算法得到的每一个块的运动矢量 $(\Delta x, \Delta y)$ 需要保存下来,作为执行第二环节运动矢量残差分级阈值法的初始化输入。

(2)使用上一步块匹配算法得到的运动矢量 $(\Delta x, \Delta y)$ 对当前帧每一个块做运动补偿,得到当前帧每一像素点的预测值 I_k 。因为 $(\Delta x, \Delta y)$ 代表的是由于物体本身运动所引起的帧间局部运动位移 MV_1 ,做帧间运动补偿以后,则可以通过阈值化当前帧像素 (i, j) 的亮度 I_k 与下一帧中 $((i, j) + MV_1)$ 位置处像素的亮度 \bar{I}_k 的差异来判断当前帧像素 (i, j) 是全局运动区域像素还是局部运动区域像素。

(3)使用阈值 C_g 分割全局运动区域与局部运动区域。阈值 C_g 的取值如式(2)所示,其中,因子 k 值的确定是本算法的关键,本文接下来通过实验统计特性来讨论因子 k 的取值。

图 1 是在采用不同的 C_g 值下,对全景视频“village”里的第十六帧的全局运动参数的估计结果。 C_g 的取值: $0 \sim 255$ 。其中,“86”代表该帧做运动估计/补偿后,由 Sobel 算子计算的平均亮度残差值;“113”代表该帧的临界阈值。

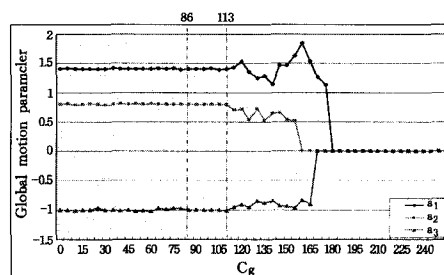


图 1 不同的 C_g 值下所得到的全局运动参数 a_1, a_2, a_3 的估计结果

由图 1 可以清楚地看到,当 C_g 的取值超过 113,全局运动参数 a_1, a_2, a_3 的估计结果精确度急剧下降,直至为零,也就是全局运动估计失败。这是由于当 C_g 的值超过临界点,过多的全局运动像素点被排除出全局运动估计过程,将导致全局运动估计结果严重偏差。另一方面,从图 1 中我们注意到,由 Sobel 算子计算的当前帧平均亮度残差值“86”与阈值临界

点“113”还有一定距离,如果单纯地使用平均亮度残差值“86”为当前帧的全局运动估计阈值,将会使得参与全局运动估计的全局运动像素集合包含过多的局部运动像素点,全局运动参数估计过程将不得不迭代多次才能逐渐排除局部运动像素点的干扰,从而得到一个较为稳定的参数结果。由以上分析得知,全景视频“village”里第十六帧的亮度阈值应取为 113,即 $C_g = 113$,则该帧的全局运动估计在确保结果准确性的前提下可以达到更快的迭代过程收敛的目的。

C_g 是一个临界点阈值。为了正确客观地估计 C_g 的值,我们对一些典型的测试序列,包括 QCIF 格式的“Bus”,“Container”,“Football”,“News”和 CIF 格式的“Mobile”,“Foreman”,“Tempete”,“Silent”,“Paris”以及全景视频序列,包括 720×576 格式的“Forest”,“Children”,“Reporter”, 1280×720 格式的“Bridge”,“Car”,“Garden”和 1920×1024 高清格式的“Village”,“Hall”,“City”做了大量的统计分析工作。我们令运动补偿后的亮度残差均值 $G_{av} = \frac{1}{N} \sum_{i=1}^N G_i$,实验结果如表 1 所列。

表 1 不同的测试序列下亮度残差均值 G_{av} 和亮度残差阈值 C_g 的取值

sequence	G_{av}	C_g	sequence	G_{av}	C_g
Mobile (CIF)	110	121	Village (1920×1024)	136	174
Foreman (CIF)	109	122	Hall (1920×1024)	131	171
Tempete (CIF)	109	119	City (1920×1024)	132	168
Silent (CIF)	102	118	Bridge (1280×720)	122	155
Paris (CIF)	108	122	Car (1280×720)	126	158
Bus (QCIF)	92	108	Garden (1280×720)	128	157
Container (QCIF)	82	93	Forest (720×576)	111	143
Football (QCIF)	91	105	Children (720×576)	118	142
News (QCIF)	86	96	Reporter (720×576)	115	141

从表 1 可以看出,亮度残差均值 G_{av} 一般情况下并没有达到临界值,如果单纯使用当前帧的亮度残差均值作为阈值点,将不能完全地从全局运动像素点集合中剔除所有的异常点。同时,也观察到,亮度残差均值 G_{av} 和阈值临界点 C_g 的值随着视频编码码率高低的的不同而不同。码率越高, G_{av} 和 C_g 的值越大,反之, G_{av} 和 C_g 的值越小。这是因为视频序列的码率越高,场景的细节更多,运动特征更明显,纹理更丰富。

根据表 1 有效的统计分析,并且考虑到阈值临界点 C_g 的取值过大将会导致全局运动参数估计失败,确定式(3)中 k 的取值为:低码率情况下,如编码 CIF, QCIF 格式的视频序列, $k=1.1$;高码率情况下,如编码 4CIF、高清、全景视频序列, $k=1.2$,也即如下式所示:

$$\begin{aligned} C_g &= 1.1G_{av}, \text{低码率情况下} \\ C_g &= 1.2G_{av}, \text{高码率情况下} \end{aligned} \quad (3)$$

3 运动矢量残差分级阈值法

3.1 运动矢量残差法

运动矢量残差分级阈值法主要是基于运动矢量残差法^[13]分离局部运动区域的过程。因为局部运动是前景物体所做的个体运动,全局运动是摄像机运动引起的背景运动,两者的差异往往非常大,根据这一准则,如果当前块的块匹配运动矢量与全局运动矢量的残差值比较大,超过了某一阈值,则可判定该当前块为局部运动区域块,反之,该当前块是全局运动区域块,如式(4)所示。

$$D = \sqrt{(\Delta x - \Delta x')^2 + (\Delta y - \Delta y')^2} \quad (4)$$

式中, $(\Delta x, \Delta y)$ 是当前块经过传统的块匹配算法得到的运动矢量, $(\Delta x', \Delta y')$ 是由当前块所选择的运动模型计算方程以及当前模型参数计算得到的该点的运动矢量。

3.2 分级阈值方法

运动矢量残差法的原理主要是基于全局运动矢量与局部运动矢量的差异,对这一差异的度量是该方法的关键点。本文采用一种分级阈值的方法来实现该度量的准确划分。这个算法的主要思路是:在迭代的初始阶段,先用一个大的百分比阈值去除绝大部分的外点。因为在一开始的时候,与内点差异较大的外点在图像中占的比例是最大的(特别是局部运动区域,往往与全局运动区域有较为明显的运动性质差异),采用一个大的百分比阈值可迅速去除这些差异性大的外点,若先排除了差异性较大的外点的干扰,则整个迭代过程收敛得更快;第一次迭代后,已经得到了一个粗略的全局运动估计结果,这个时候图像中剩下的待处理的外点是与内点差异较小的像素点,在图像中占的比例不大,在接下来的几次迭代中,可以分级减少百分比阈值的值,逐渐排除这些差异性小的外点,精确细化内点的集合,直到把尽可能多的外点都排除出内点集合。具体算法步骤如下:

(1)使用块匹配算法估计当前帧每一块的局部运动位移。该步骤在上一环节,即基于亮度残差的阈值法中已经执行。

(2)估计运动模型参数。考虑到全局运动区域一般集中在图像背景区域,也就是图像的边缘部分,我们选取图像左上角、右上角、左下角 3 个 48×48 图像区域的中心块(块大小为 16×16)的运动矢量作为第一次高斯牛顿迭代估计模型参数过程的初始化输入值,通过高斯牛顿迭代计算,可得出初始化的若干个运动模型参数值。

(3)计算图像中每一个块的全局运动矢量 $(\Delta x', \Delta y')$ 。图像中,背景区域的块运动属于全局运动,而前景区域的块运动属于局部运动,但仍然可以通过把上一步骤中得到的运动模型参数值重新代入高斯牛顿迭代中,根据每一个块中心像素点的坐标 (x, y) 的值,分别计算出每一个块的“全局运动矢量”。对于局部运动块,该运动矢量是伪全局运动矢量。

(4)计算运动矢量残差。对于同一个块,把步骤(1)得出的该块的运动矢量 $(\Delta x, \Delta y)$ 与步骤(3)中得出的该块的运动矢量 $(\Delta x', \Delta y')$ 作方差,如式(4)所示。计算运动矢量残差的目的是为了判断该像素块的实际运动矢量与全局运动矢量的差异,如果运动矢量残差较大,则说明该块很有可能是局部运动块,反之,该块很有可能是全局运动块。

(5)设置百分比阈值来排除局部运动块。我们采用一个分级递减的百分比阈值 P 来逐步精细全局运动区域,同时设置一个分级递减的运动矢量残差阈值 T 来判断本次迭代是否可以结束。在第一次迭代过程中,设置 P 的初始值为 50,同时设置 T 为 0.04,即:

$$\sqrt{(\Delta x - \Delta x')^2 + (\Delta y - \Delta y')^2} > T \quad (5)$$

把满足上式条件的最多不超过图像块数目总和的 P 百分比的块标记为局部运动块,剩余的标记为全局运动块,作为下一次全局运动估计迭代计算的块集合。如果本次全局运动估计块集合里所有块的运动矢量残差都不满足上式条件,则结束本次迭代,生成最终的运动模型参数,同时可根据每次迭代所标记的局部运动块集合分割出局部运动区域。 P 和 T 的递减规则:每次进入新一轮的迭代计算前,将 P 更新为 P 的四分之一($P = P/4$),即第二次迭代过程中, P 设为 12.5,第

三次迭代过程中, P 设为 3.125……这是因为随着迭代次数的增加, 内点集合越来越趋于稳定, 所包含的外点越来越少, 所以 P 值需要线性递减, 以免迭代收敛于内点集合的局部。同时将 T 更新为 $T=T-0.01$, 因为每次迭代结束以后, 剩下的块的运动矢量残差更小, 其中有一部分是与全局运动属性差异更小的局部运动块, 需要用更小的 T 来分辨它们。当迭代结果趋向稳定时, 用递减的 T 能同时达到加快迭代收敛、减低计算复杂度和保证全局运动估计结果逐步精细的目的。第一次高斯牛顿迭代结束以后, 将得到新的一套运动模型参数值。把这套运动模型参数值作为下一轮高斯牛顿迭代的输入值。

(6) 重复步骤(3)一步骤(5), 直到所有块的运动矢量残差都小于 T 或者迭代次数等于 4, 则结束迭代过程。实验证明, 当迭代次数超过 4 次, 模型参数的解的精确性增加极其有限, 也就是说, 第 4、第 5、第 6 次迭代的结果非常近似, 甚至无异, 模型参数的解已经达到了该分级阈值迭代法精确性的一个极限。

4 实验结果与分析

在本文仿真实验中, 采用六个视频序列, 涵盖从高分辨率到低分辨率, 包括分辨率为 2048×768 的“Village”和“Hall”, 分辨率为 1280×352 的“City”和“Bridge”以及分辨率为 720×176 的“Car”和“Garden”。帧率都是每秒 25 帧。

如上面讨论可知, 运动模型参数估计过程是全局运动估计中较为耗时的部分, 迭代次数越多, 计算复杂度越高。而迭代收敛的快慢主要受全局运动像素点集合的纯度的影响, 全局运动像素点集合纯度越高, 迭代收敛得越快; 反之, 越慢。如图 2 所示, 采用传统的基于平移运动模型的运动估计/运动补偿方法, 平均需要 5~6 次高斯牛顿迭代才能趋于全局运动参数的精确解; 而采用本文所提出的渐进精细阈值法的快速全局运动估计技术, 平均使用 2~3 次高斯牛顿迭代就可获得精度比前者更高的全局运动参数解。实验证明, 在运动剧烈、运动纹理更丰富的视频序列, 例如“Bridge”, 采用本文所提出的方法能更有效地提高运动预测精度和编码性能。而运动剧烈的视频序列往往是传统运动估计/运动补偿难以处理的视频内容。

在已有的全局运动模型参数估计过程中亮度残差阈值法和运动矢量残差分级阈值法并不会给编码器带来很大的额外计算开销。表 2 列出了本文所提出的快速全局运动估计方法中各步骤所需要的计算运行时间。块匹配过程和全局运动参数估计过程是全局运动估计过程中计算复杂度开销最大的部

分, 平均占总运行时域分别为 50.6% 和 32.5%, 而执行亮度残差阈值法和运动矢量残差分级阈值法二者加起来的运算时间不到总计算开销的 17%。

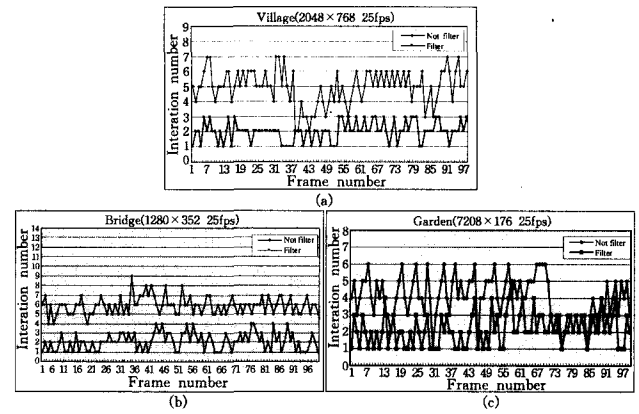


图 2 采用阈值方法与不使用阈值方法两种情况下全局运动参数估计过程的迭代次数比较

表 2 快速全局运动估计各步骤运算时域

Sequence	Block match (ms)	Improved average threshold (ms)	Global motion parameter estimation (ms)	Hierarchical threshold (ms)	Total time (ms)
Village	1172 (54.6%)	112 (5.2%)	721 (33.6%)	140 (6.6%)	2145
Hall	926 (47.3%)	187 (9.6%)	632 (32.3%)	211 (10.8%)	1956
City	705 (46.2%)	128 (8.4%)	565 (37%)	126 (8.4%)	1524
Bridge	852 (52.9%)	101 (6.3%)	526 (32.6%)	133 (8.2%)	1612
Car	483 (50.2%)	85 (8.8%)	284 (29.5%)	111 (11.5%)	963
Garden	535 (52.3%)	82 (8%)	306 (29.9%)	100 (9.8%)	1023
Average	50.6%	7.7%	32.5%	9.2%	100%

通过加入阈值器, 以较小的计算开销代价, 使得全局运动参数估计过程收敛得更快, 全局运动估计结果更精确。表 3 列出了渐进精细阈值法的全局运动估计方法与其他 3 种分别基于固定值阈值法、均值阈值法、固定百分比阈值法的全局运动估计方法的全局运动估计时域和重构图像峰值信噪比的对比数据。本文所采用的快速全局运动估计方法在保证重构图像质量不下降, 而且还有略微提升的前提下, 分别比固定百分比阈值法节省 25.6%, 比均值阈值法节省 32.5%, 比固定值阈值法节省 52.3% 的运动估计计算开销。

表 3 使用渐进精细阈值法与其他阈值法的性能比较

Sequence	Proposed method		Constant percentage threshold method		Average threshold method		Constant threshold method	
	GME time(ms)	PSNR(dB)	GME time(ms)	PSNR(dB)	GME time(ms)	PSNR(dB)	GME time(ms)	PSNR(dB)
Village	2145	32.29	2542	32.26	2632	32.22	4021	32.16
Hall	1956	32.46	2216	32.43	2516	32.42	4184	32.33
City	1524	33.1	2084	33.05	2265	33.06	2753	33.02
Bridge	1612	32.08	1968	32.03	2314	32.05	3335	32.02
Car	963	33.2	1254	33.16	1410	33.14	1868	33.09
Garden	1023	33.75	1321	33.65	1662	33.56	1964	33.51

结束语 本文研究了一种面向快速全局运动估计的渐进精细阈值方法。该方法结合亮度残差阈值法和运动矢量残差阈值法, 分两步逐步细分前景物体对象和背景区域, 最终实现快速全局运动估计。实验证明, 该阈值方法兼顾了全局运动估计的准确性和鲁棒性, 对大多数视频序列的全局运动估计

具有算法复杂度低、运动参数估计精确的效果。

通过渐进精细阈值法, 可以得到全局运动的模型参数、全局运动区域和局部运动区域。从理论上说, 全局运动区域对应了视频场景中的背景区域, 而前景运动目标对应了局部运动区域, 但是由于块匹配估计运动矢量时的孔径问题和噪声

问题,全局运动区域中还存在一些前景运动物体像素。我们下一步的研究工作是通过全局运动区域的一些辅助处理,如运动矢量场的后处理、时空联合分析等措施来进一步净化全局运动区域,从而得到更为精确的全局运动估计结果。

参考文献

[1] 郭丽,龚声蓉,崔志明.一种快速全局运动估计算法[J].中国图形图像学报,2007,12(12):2119-2126

[2] 李乔亮,汪国有,张桂林,等.基于掩模金字塔的高精度全局运动估计算法[J].计算机辅助设计与图形学学报,2009,21(6):758-762

[3] Arvanitidou M G, Glantz A, Krutz A, et al. Global motion estimation using variable block sizes and its application to object segmentation[C]//Proceedings of Image Analysis for Multimedia Interactive Services. 2009:173-176

[4] 彭小宁,杨明,邹北骥,等.基于局部图金字塔的不规则块匹配视频分割方法[J].计算机科学,2008,4:233-237

[5] Zheng Jia-li, Ji Xiang-yang, Ni Guang-nan, et al. Extended Direct Mode for Hierarchical B Picture Coding [J]. Proceedings of IEEE International Conference on Image Processing (ICIP), 2005,2(1):265-268

[6] Press W H, et al. Numerical Recipes in C: the Art of Scientific Computing[M]. Cambridge: Cambridge University Press, 2005

[7] Zheng Jia-li, Shen Yan-fei, Zhang Yong-dong, et al. Adaptive Selection Motion Model for Panoramic Video Coding[C]// Proceedings of IEEE International Conference on Multimedia & Expo (ICME), 2007:1319-1322

[8] Rath G B, Makur A. Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 9(7): 1075-1099

[9] Li W P, et al. MPEG-4 Video Verification Model version 16. 0, ISO/IEC/JTC1/SC29/WG11, Doc. No. N3312[S]. Netherlands, 2006

[10] Albu F, Florea C, Zamfir A, et al. Low Complexity Global Motion Estimation Techniques for Image Stabilization[C]// Proceedings of International Conference on Consumer Electronics, 2008:1-2

[11] Bin Q, Ghazal M, Amer A. Robust Global Motion Estimation Oriented to Video Object Segmentation[J]. IEEE Transactions on Image Processing, 2008,17(6):958-967

[12] Gonzales R C, Woods R E. Digital Image Processing[M]. Addison-Wisley, 2005

[13] Chen Y M, Bajic I V. Motion Vector Outlier Rejection Cascade for Global Motion Estimation[J]. IEEE Transactions on Signal Processing, 2009,17(2):197-200

(上接第 269 页)

本文定义了“进球”、“点球”和“角球”3类视频语义图。实验数据如表2所列。

表2 实验数据集

序号	比赛名称	长度	机构	时间
1	德国 VS 波兰	1 小时 35 分	CCTV5	2006
2	瑞典 VS 巴拉圭	1 小时 37 分	CCTV5	2006
3	意大利 VS 美国	1 小时 33 分	东方卫视	2006
4	巴西 VS 日本	1 小时 34 分	东方卫视	2006

实验结果如表3所列,从表中可以看出查找到率都高于查准率,而且平均准确率大于85%。这说明采用VSG描述的五类足球视频语义是准确的,且本文提出的算法实用可行。

表3 实验结果

足球语义	实际包含数量	正确匹配数量	错误匹配数量	查准率 (%)	查到率 (%)
进球	10	10	3	76.9	100
角球	40	34	0	85	85
点球	2	2	0	100	100
平均值				87.3	95

分析实验结果发现,造成“进球”匹配错误主要是由视频中出现的精彩射门引起的。而“角球”的正确匹配数量小于实际数量的原因是比赛视频中,有些角球镜头没有给出角球区镜头。相比之下,“点球”的匹配准确率最高,这是因为点球的情况较少,并且描述点球的视频一般比较规范,受其他情况影响较少。

采用简单线性加权的融合方法^[4]对测试数据进行实验,查准率和查全率的平均值分别是42%和83%,可见本文的方法明显优于基于简单线性加权的融合方法。

结束语 提出了新的图模型的视频语义内容建模方法,构建了视频语义图(VSG)。从3个方面详细地阐述了图模型,第一,如何定义每个建模元素的图形符号;第二,如何定义

图形符号之间的位置关系;第三,如何将表示法映射到抽象语义。针对足球视频,对模型进行了深入的讨论,提出了基于图模型的足球视频语义内容建模方法。可以将其推广到其他视频语义内容建模,允许用户自行定义图形元素表示的语义。在分析了图匹配的相关问题之后,提出了基于DFS的图模型的视频检索算法,详细阐述了算法思路和步骤,分析了算法的复杂度,并用实验验证了算法的可行性。

参考文献

[1] Jain R. Multimedia Information Retrieval; Watershed Events. MIR'08[C]// Vancouver, British Columbia, Canada, October 2008:30-31

[2] Rui Y, Gupta A, Acero A. Automatically extracting highlights for TV baseball programs[C]// ACM Multimedia, 2000

[3] Wang Feng, Jiang Yu-gang, Ngo C-W. Video Event Detection Using Motion Relativity and Visual Relatedness[C]// ACM MM'08. Vancouver, British Columbia, Canada, 2008

[4] 白亮. 本体支持的视频情报分析方法与技术研究[D]. 长沙:国防科技大学, 2008

[5] 李刚. 知识发现的图模型方法[D]. 北京:中国科学院软件研究所, 2001

[6] 严蔚敏,吴伟明. 数据结构[M]. C语言版. 北京:清华大学出版社, 1997

[7] Goldberg A B, Li M, Zhu X. Online manifold regularization: A new learning setting and empirical study[C]// ECML PKDD, 2008

[8] Aly R, Hiemstra D. Concept Detectors: How Good is Good Enough[C]// MM'09. Beijing, 2009

[9] Han Bo, Yan Yan, Chen Zhen-ghua, et al. A General Framework for Automatic On-line Replay Detection in Sports Video[C]// MM'09. Beijing, 2009

[10] 侯绿林,白亮,老松杨.一种压缩域中的体育视频慢镜头探测方法[J].计算机科学,2009,36(9):283-286