

基于视觉运动特性的视频时空显著性区域提取方法

周 莺^{1,2} 张基宏¹ 梁永生² 柳 伟²

(深圳大学信息工程学院 深圳 518060)¹

(深圳信息职业技术学院可视媒体处理与传输深圳市重点实验室 深圳 518172)²

摘要 为了更准确有效地提取人眼观察视频的显著性区域,提出一种基于视觉运动特性的视频时空显著性区域提取方法。该方法首先通过分析视频每帧的频域对数谱得到空域显著图,利用全局运动估计和块匹配得到时域显著图,再结合人眼观察视频时的视觉特性,根据对不同运动特性视频的主观感知,动态融合时空显著图。实验分析从主客观两个方面衡量。视觉观测和量化指标均表明,与其他经典方法相比,所提方法提取的显著性区域能够更准确地反映人眼的视觉注视区域。

关键词 显著性区域,视觉注意模型,时域显著度,空域显著度,运动特性

中图分类号 TP751 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.11.025

Motion Characteristics Based Video Salient Region Extraction Method

ZHOU Ying^{1,2} ZHANG Ji-hong¹ LIANG Yong-sheng² LIU Wei²

(College of Information Engineering, Shenzhen University, Shenzhen 518060, China)¹

(Shenzhen Key Laboratory of Visual Media Processing and Transmission, Shenzhen Institute of Information Technology, Shenzhen 518172, China)²

Abstract The human eyes only observe the salient region of the video. Thus a motion characteristics based salient region extraction method was proposed. Spatial saliency map is extracted by analyzing the log spectrum of each frame in the frequency domain. Temporal saliency map is obtained by global motion estimation and block matching. According to the human visual characteristics and the subjective perception of different motion characteristics, the region of saliency is fused dynamically by spatial and temporal saliency map. The experiment was analyzed from both subjective and objective indicators. Visual observation and quantitative indicators show that the proposed method can reflect the human visual attention area more accurately than other classical extraction methods.

Keywords Region of saliency, Visual attention model, Temporal saliency, Spatial saliency, Motion characteristics

1 引言

人眼视觉系统(Human Visual System, HVS)观察一段视频或图像时,通常进入视野的视觉信息量是很大的,这些信息无法同时被人脑接收和处理,并且这些信息的重要度也不同。而HVS具有一定的选择性,只对画面中的部分内容或区域感兴趣,这些能显著吸引人眼注意力的区域称为显著性区域(Region of Saliency, ROS),ROS的存在是HVS心理特征的重要表现^[1];其他区域则称为非显著性区域或背景区域。

显著性区域在目标识别与跟踪、视频压缩编码、场景分类、图像检索等领域有着广泛的应用。在如何高效、准确地获取视频显著性区域方面,国内外学者进行了大量研究工作。本文主要结合HVS对于视频的视觉感知特性,提出一种融合时空域的显著性区域提取方法。

2 视觉注意机制

人们观察复杂场景时,迅速将注意力集中在少数重要区域,并利用有限的处理能力对其优先处理^[2],这就是视觉注意机制。利用视觉注意模型(Visual Attention Model, VAM)提取视频显著性区域,更加符合人眼的生理特性,是一种较为准确的提取方法。视觉注意模型包括自顶向下(Top-down)的任务驱动型和自底向上(Bottom-up)的数据驱动型两种实现方式。前者结合主观认知及视觉场景分析,大多采用机器学习的方法,根据视觉特征统计数据或先验知识来提取显著性区域;后者通过模拟人类的视觉、知觉刺激过程,提取场景的不同特征(如纹理、颜色、方向、密度等)作为图像的显著度,对其建模并选出显著度大的注意区域。本文提取视频显著性区域也是基于自底向上的视觉注意模型。

最典型的自底向上VAM是1998年美国加州大学的Itti

到稿日期:2014-07-17 返修日期:2014-11-18 本文受国家自然科学基金项目(61172165),广东省自然科学基金项目(S2011010006113)资助。

周 莺(1982-),女,博士生,主要研究方向为多媒体通信、信号处理、网络仿真,E-mail:zhouying722@163.com;张基宏(1964-),男,教授,博士生导师,主要研究方向为图像处理、神经网络、多媒体通信;梁永生(1971-),男,教授,主要研究方向为计算机网络与数据通信、信号处理与模式识别。

提出的。Itti 模型^[3]提取图像的亮度、颜色和方向特征,进行多尺度融合得到显著图。Harel 提出了另一种类似的 GBVS 模型^[4](Graph Based Visual Saliency),该模型在特征提取之后,计算活动图并进行归一化得到显著图。Walther 在 Itti 模型的基础上引入神经网络中的竞争机制检测显著对象^[5]。伦敦大学的 Stentiford^[6]提出一种基于上下文的模型,使用随机邻域概念,在图像剩余区域取其邻居像素与随机抽取的其他若干个像素作比较,从中得到显著图。文献^[7]基于傅立叶变换,在频域利用灰度特征,对剩余谱和输入图像的相位谱做傅立叶逆变换得到显著图。以上视觉注意模型在提取单幅图像的显著图方面都取得了不错的效果,然而,将其应用在视频的显著性区域提取时,由于未考虑视频帧之间的相关性以及人眼观察视频的视觉特性,因此提取的显著图与人眼观察的显著性区域存在较大差异。

也有一些方法综合考虑了视频的空域和时域特性。Itti 通过增加运动特征和闪烁将静态模型扩展到视频中。文献^[8]提出了一种融合运动和空间关系特性的显著性检测方法,在颜色对比度的基础上,引入图像中目标的空间深度关系和运动特征。文献^[9]融合了时域相关性和空域显著度。Fang 等人^[10]在压缩域利用离散余弦变换提取 I 帧的亮度、颜色、纹理特征以及 P、B 帧的运动特征加以融合得到显著图。

事实证明,人眼在观看视频时,对于运动变化的物体或者与周围反差较大的对象反应敏感。本文根据 HVS 的这一视觉感知特性,提取能够反映人眼视觉特性的空域和时域特征,根据不同内容视频的动态性加权融合得到显著性区域。实验结果表明,利用本文方法提取的显著性区域更加符合人眼对视频内容的主观感知。

3 基于运动特性的时空显著区域提取

3.1 空域显著度计算

Hou 指出^[11],HVS 发现并确定显著区域可以分为两步:首先是并行、快速并且简单的“预注意”阶段,该阶段通过整合某些初级特征(如方向、边界、密度等)锁定一些主要目标,这些主要目标作为候选,再经过第二阶段认真、缓慢并且复杂的“注意”处理,从中确定显著区域。然而,从图像或视频中选取有代表性的特征并建立视觉注意模型是非常复杂的过程,特征选取的好坏,以及模型设计的复杂度,将直接影响显著区域提取的准确度。

根据信息论,图像在压缩编码时由两部分组成,一部分是新颖的代表图像特征的信息,一部分是优先获取的信息,后者在编码过程中可以作为冗余信息进行压缩。大多数图像的频域对数谱符合一种简单的线性关系,因此,本文采用文献^[11]的方法,通过分析并处理频域内图像的对数谱,去除图像的大部分冗余信息,再经过傅立叶反变换,便可得到图像的空域显著图。具体算法如式(1)~式(3)所示:

$$A(f) = \log |\hat{F}(f)| \quad (1)$$

$$R(f) = A(f) - I(A(f)) \quad (2)$$

$$S_s(f) = g(f) * \hat{F}^{-1}[\exp(R(f) + P(f))]^2 \quad (3)$$

假设输入图像为 f ,则 $\hat{F}(f)$ 为 f 的傅立叶变换; $A(f)$ 为 f 的对数振幅谱; $I(A(f))$ 为 $A(f)$ 滤波后的函数输出, $R(f)$

为 f 的频谱冗余; $P(f)$ 为 f 的相位谱, $g(f)$ 为对 f 进行高斯滤波, $S_s(f)$ 即输出的空域显著图。

利用上述方法分别对 H. 264/SVC 标准测试序列中的 News、Mobile 和 Soccer 序列(CIF)中的某一帧提取空域显著图,结果如图 1 所示。



图 1 视频序列当前帧空域显著图

从图 1 的结果可以看出,基于频谱的显著图提取方法能够提取图像中的大部分显著性区域。

3.2 时域显著度计算

视频与图像的不同之处在于它包含运动信息。神经心理学研究表明,HVS 观察视频时,在视觉注意机制的作用下,更容易被运动变化的物体或场景所吸引,而且人眼对具有不同运动特性的视频的主观感知是不一样的。

实际的运动场景主要有以下 3 种^[12]:1)背景无变化或变化不大,只有前景对象在变化。观察这类视频时,人眼主要关注运动变化的对象。2)背景运动剧烈。若该帧运动矢量的总长度很大,则人眼更注重相对变化较小的对象;若大部分物体变化不是非常剧烈,则人眼对高速运动的部分较敏感。3)整个场景都剧烈运动,则人眼很难分辨视频中的细节,而更关心场景中运动不明显的对象。

本文中,通过视频的运动特性来描述人眼对于不同视频内容变化的敏感程度。一段视频中,由摄像机运动产生的像素运动称为全局运动,目标对象的运动称为局部运动。人眼主要对目标对象的运动变化感兴趣,因此要描述视频内容的运动特性,首先需要根据全局运动估计模型区分前、背景的不同运动信息并提取目标对象。

本文采用六参数模型估计摄像机的全局运动,该模型可对复杂的几何变换运动进行建模,如式(4)所示:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \quad (4)$$

其中, $(x' \ y')$ 为当前块中心相对于图像中心的坐标, $(x \ y)$ 为参考帧中匹配块在当前帧相同位置的坐标; a_1, a_2, a_3, a_4 决定缩放与旋转; a_5, a_6 分别决定水平和垂直移动。定义样本点集合的误差函数如式(5)所示:

$$E(R) = \sum_{i=1}^k [(a_1 x_i - a_2 y_i + a_5 - x_i')^2 + (a_3 x_i + a_4 y_i + a_6 - y_i')^2] \quad (5)$$

根据最小化误差函数的准则,即

$$R_{opt} = \min_R E(R) \quad (6)$$

通过迭代最小二乘法,可求得最优全局运动估计参数,如式(7)、式(8)所示:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_5 \end{pmatrix} = \begin{bmatrix} \sum_{i=1}^k x_i^2 & \sum_{i=1}^k x_i y_i & \sum_{i=1}^k x_i \\ \sum_{i=1}^k x_i y_i & \sum_{i=1}^k y_i^2 & \sum_{i=1}^k y_i \\ \sum_{i=1}^k x_i & \sum_{i=1}^k y_i & k \end{bmatrix} * \begin{pmatrix} \sum_{i=1}^k x_i x_i' \\ \sum_{i=1}^k y_i x_i' \\ \sum_{i=1}^k x_i x_i' \end{pmatrix} \quad (7)$$

$$\begin{pmatrix} a_3 \\ a_4 \\ a_6 \end{pmatrix} = \begin{bmatrix} \sum_{i=1}^k x_i^2 & \sum_{i=1}^k x_i y_i & \sum_{i=1}^k x_i \\ \sum_{i=1}^k x_i y_i & \sum_{i=1}^k y_i^2 & \sum_{i=1}^k y_i \\ \sum_{i=1}^k x_i & \sum_{i=1}^k y_i & k \end{bmatrix} * \begin{pmatrix} \sum_{i=1}^k x_i y_i' \\ \sum_{i=1}^k y_i y_i' \\ \sum_{i=1}^k x_i y_i' \end{pmatrix} \quad (8)$$

利用块匹配运动估计算法获取当前帧的运动矢量,得到全局运动估计参数后,在参考帧和当前帧之间进行一次背景映射,便可提取出运动目标。

对目标对象的运动矢量图进行高斯滤波,再利用视觉注意模型,便可得到视频当前帧的时域显著图。其算法流程如图2所示。

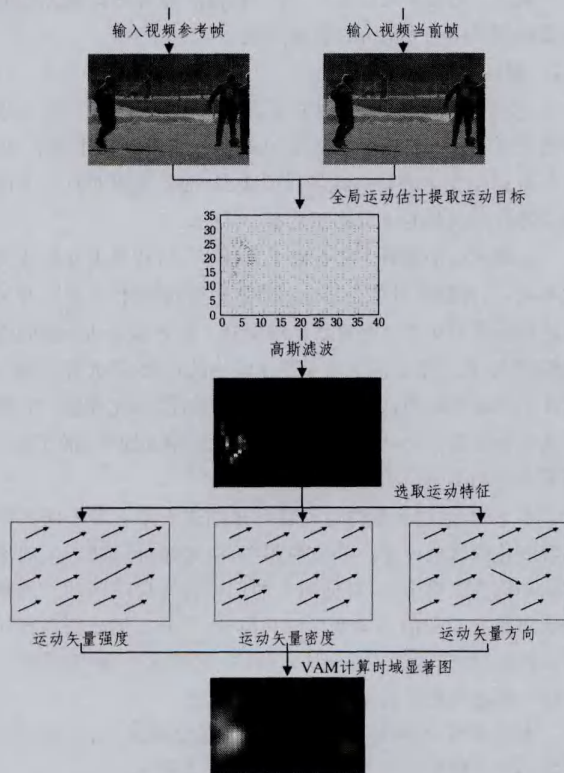


图2 视频序列当前帧时域显著图提取流程

3.3 基于运动特性的时空显著区域提取

根据HVS对视频场景的主观感知特性,本文提出一种基于运动特性的时空显著区域提取方法。该方法在分别提取反映视频结构的空域显著图和反映视频运动特性的时域显著图之后,采用线性加权组合的方式,根据视频运动特性自适应

融合时空显著图,如式(9)所示:

$$S = \frac{\sum_{i=1}^N \omega_S \times S_S(i) + \sum_{i=1}^N \omega_T \times S_T(i)}{\sum_{i=1}^N \omega_S + \sum_{i=1}^N \omega_T} \quad (9)$$

其中, $S_S(i)$ 表示视频序列第*i*帧的空域显著图, $S_T(i)$ 表示时域显著图, S 即为融合后的时空显著图。 ω_S, ω_T 为融合权重。

根据人眼的向心性及其显著区域分布与人眼关注度的关系^[13],假设图像中心点位置为 (x, y) ,显著区域任一点像素位置为 (x_i, y_i) ,则:

$$\omega_S = \| (x_i, y_i) - (x, y) \| * \frac{1}{2\pi\sigma} \exp\left(-\frac{(x_i-x)^2 + (y_i-y)^2}{2\sigma^2}\right) \quad (10)$$

式中, ω_S 为高斯加权的欧氏距离,显著区域离中心点越近,分布越集中,则人眼对其越敏感,显著性权重越高。

ω_T 则根据视频运动特性动态调节,同时考虑运动的空域分布 ω_1 、运动强度 ω_2 和运动复杂度 ω_3 这3个因素,令

$$\omega_T = \omega_1 \times \omega_2 \times \omega_3 \quad (11)$$

$$\omega_1 = \frac{N_O}{N(s)} \quad (12)$$

式(12)中, N_O 为运动目标非零运动矢量的宏块数, $N(s)$ 为该帧的宏块数, ω_1 即为运动空域分布值。

$$\omega_2 = \frac{\sum_{i=1}^{N(s)} (\|v_x\| + \|v_y\|)}{\sum_{i=1}^{N(s)} (\|v_x\| + \|v_y\|)} \quad (13)$$

式中, v_x 和 v_y 分别为目标对象运动矢量的横、纵坐标, ω_2 为运动能量的大小,其值越大说明运动信息越丰富。

$$\omega_3 = \frac{-[\sum_{i=1}^m \frac{N(s_i)}{N(s)} \times \log(\frac{N(s_i)}{N(s)})]}{\lg(36)} \quad (14)$$

式中, s_i 是目标对象运动矢量方向直方图中各非空的维度; $N(s_i)$ 是各个维度内运动矢量非零的宏块数, $i \leq 36$,通过信息熵求得各宏块运动矢量在各维度的分布。

当视频所含的运动信息较为丰富并占有优势时,时域显著图的权重将有所增加;反之,则减小时域显著图的权重。

4 实验结果与分析

为了验证本文提出的基于运动特性的时空显著区域提取方法更加符合HVS观察视频时的主观感受,将利用本文算法提取的显著区域与分别采用Itti^[3], Harel^[4], Walther^[5], Hou^[7]4种VAM提取的显著区域进行对比。测试视频为H.264/SVC标准测试序列中的Walk、Foreman、Tempete和Soccer4个序列(CIF),并从主、客观两个方面对实验结果进行评价。

首先利用SMI眼动仪获取人眼在观察视频时可能注视的区域,构造客观注意图(Ground Truth),获取Ground Truth后,将利用各算法提取的显著图与Ground Truth进行比较,结果如图3、图4所示。

对比图3、图4中提取的显著图与Ground Truth可以看出,本文提出的方法较其他方法提取的显著图更加符合人眼的主观感知。

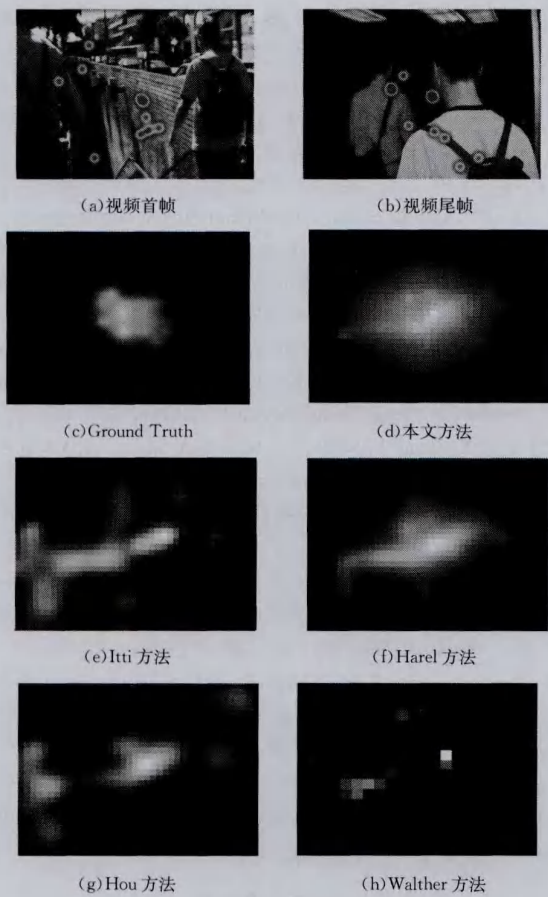


图3 Walk序列显著图比较

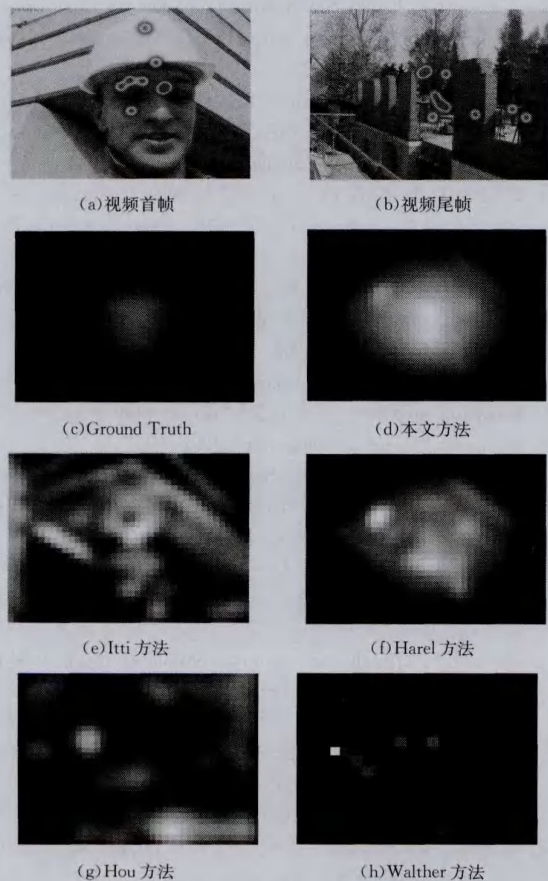


图4 Foreman序列显著图比较

ROC(Receiver Operating Characteristic)作为衡量二值分类器在不同判断标准下的效果,常被用来作为评价视觉效果指标。由预测的显著对象图和 Ground Truth 显著度图的眼动数据可得:

真阳性率=(预测为显著对象 & 实际为注视点)/((预测为显著对象 & 实际为注视点)+(预测为非显著对象 & 实际为注视点))

假阳性率=(预测为显著对象 & 实际为非注视点)/((预测为显著对象 & 实际为非注视点)+(预测为非显著对象 & 实际为非注视点))

在性能比较上,本文采用 ROC 曲线及 ROC 曲线下的面积 AUC(Area Under the ROC Curve)值来评价显著图提取的好坏。AUC 值越大,说明算法性能越好。各算法对实验视频序列的 ROC 曲线如图 5 所示,AUC 值如表 1 所列。

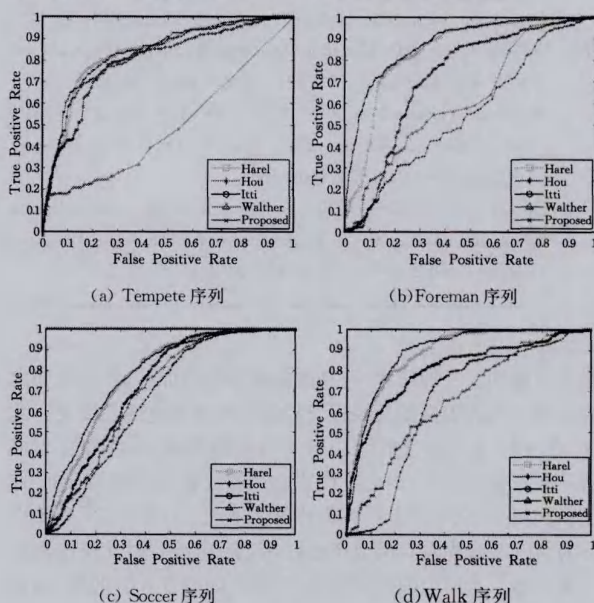


图5 各算法 ROC 曲线

表1 各算法 AUC 值

算法	视频序列			
	Tempete	Foreman	Soccer	Walk
Itti	0.7971	0.7203	0.7326	0.8105
Harel	0.8343	0.8458	0.7808	0.8873
Hou	0.8234	0.5544	0.6733	0.6543
Walther	0.4908	0.6256	0.7077	0.6647
本文算法	0.8395	0.8801	0.7926	0.8912

由图 5 和表 1 的结果可以看出,由于考虑了 HVS 对视频运动特性的主观感知,本文提出的基于运动特性的时空显著区域提取方法能够更为准确有效地确定实际人眼的注视区域。对于运动目标与背景区域反差较大的视频序列(如 Foreman、Soccer),效果尤其明显。

结束语 HVS 在观察视频时,只对其中运动变化的物体或与周围反差较大的区域感兴趣,而且对于不同运动特性的视频,人眼的主观感知也不同。本文针对 HVS 的这一视觉特性,提出一种基于视频运动特性的时空显著性区域提取方法,该方法根据视频的运动特性(强度、密度和复杂度)动态融合时域显著图和空域显著图。实验结果表明,本文提出的方法较其他经典方法能提取更加符合人眼的视觉特性和主观感知的显著图。

参考文献

- [1] 张菁,卓力,李晓光. 新一代高效视频编码技术[M]. 北京:人民邮电出版社,2013
Zhang Jing, Zhuo Li, Li Xiao-guang. High efficiency Video Coding Techniques for Next Generation [M]. Beijing: Posts & Telecom Press, 2013
- [2] 王岩. 视觉注意模型的研究与应用[D]. 上海:上海交通大学, 2012
Wang Yan. Research and Application of Visual Attention Model [D]. Shanghai, Jiaotong University, 2012
- [3] Koch C, Itti L, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259
- [4] Koch C, Harel J, Perona P. Graph-Based Visual Saliency[M]// Advances in Neural Information Processing Systems. 2007: 681-688
- [5] Walther D, Koch C. Modeling Attention to Salient Proto-objects [J]. Neural Networks, 2006, 19(9):1395-1407
- [6] Bamidele A, Stentiford F W M. An attention based similarity measure used to identify image clusters[OL]. <http://www.doc88.com/P-9592796942640.html>
- [7] Hou X, Harel J, Koch C. Image Signature: Highlighting Sparse Salient Regions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1):194-201
- [8] 刘晓辉,金志刚,赵安安,等. 融合运动和空间关系特性的显著性区域检测[J]. 华中科技大学学报, 2013, 41(6):45-49
Liu Xiao-hui, Jin Zhi-gang, Zhao An-an, et al. Salient region detection of interfusing motion and spatial relationships[J]. Journal of Huazhong University of Science and Technology, 2013, 41(6):45-49
- [9] Luo Y, Tian Q. Spatio-temporal enhanced sparse feature selection for video saliency estimation[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). 2012:33-38
- [10] Fang Y, Lin W, Chen Z, et al. A Video Saliency Detection Model in Compressed Domain[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(1):27-38
- [11] Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach[M]. Minneapolis, MN, USA, 2007
- [12] 周莺,柳伟,张基宏. 基于内容感知的可分级视频码流排序方法[J]. 信号处理, 2013, 29(8):1012-1018
Zhou Ying, Liu Wei, Zhang Ji-hong. Content aware based sorting approach of scalable video bitstream[J]. Journal of Signal Processing, 2013, 29(8):1012-1018
- [13] Gopalakrishnan V, Hu Yi-qun, Rajan D. Saliency region detection by modeling distributions of color and orientation [J]. IEEE transactions on multimedia, 2011, 2(5):892-905

(上接第 95 页)

是没有达到线性加速比,这是因为线程数增加时,各线程和内存之间交换的数据量及线程管理的时间开销也随之增大。此外,在 MIC 卡上设置的线程数并不是越多越好,线程数太多将导致开销比较大,因此要设置合适的线程数来确保 MIC 核的高利用率。例如图 4 中基于 MIC 产生 100000000 个随机数时,在 32 线程左右的效率最好,多于 32 线程时加速比出现下滑趋势。当产生随机数数量相同时,随着线程数的增加,加速比增大。

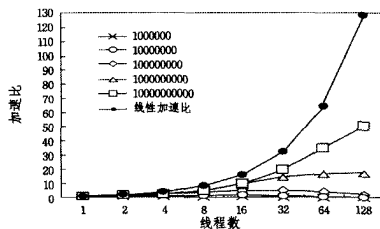


图 4 MIC 平台下的加速比

比较 CPU 和 MIC 的测试数据后发现, MIC 的速度没有 CPU 的快,这与存储器访问有关。现代处理器速度的快速发展和存储器速度的慢速发展导致处理器要花费大量的时间等待存储器数据返回。在并行计算体系结构中,计算速度更快,等待存储器数据返回的时间更长;而 MIC 众核处理器需要开启比 CPU 更多的线程,因此 MIC 上内存读写数据和等待存储器数据返回的时间更长;且 MIC 单核的性能低于 CPU,从而导致 MIC 速度低于 CPU 速度。

结束语 本文研究了基于 MIC 平台的 GFSR(521,32)并行化方法,并对 GFSR(521,32)的并行化程序进行了 TestU01 测试、编译选项优化和时间测试。并行化后 GFSR(521,32)的 TestU01 测试结果与串行程序的测试结果一致。时间测试的结果表明,当数量级较小时,加速比并不明显;随着数量级的不断增大,加速比逐渐提高。但线程数太多时线程管理

的开销较大,加速比增长反而较小。相比于 CPU 单线程, MIC 平台的最优加速比为 7.58 倍。

参考文献

- [1] L'Ecuyer P, Tezuka S. Structural properties for two classes of combined random number generators[J]. Mathematics of Computation, 1991, 57(196):742-743
- [2] L'Ecuyer P, Blouin F, Couture R. A search for good multiple recursive random number generators[J]. ACM Transactions on Modeling and Computer Simulation, 1993, 3(2):87-98
- [3] Bradley T, du Toit, Giles J, et al. Parallelization techniques for random number generators[M]// GPU Computing Gems Emerald Edition. 2011:231-246
- [4] Makino J, Miyamura O. Generation of Shift Register Random Numbers on Vector Processor[J]. Computer Physics Communication, 1991, 64(23):363-368
- [5] Makino J, Takaishi T, Miyamura O. Generation of Shift Register Random Numbers on Distributed Memory Multiprocessors[J]. Computer Physics Communication, 1992, 70(3):495-500
- [6] Wei Gong-yi, Yang Zi-qiang. Some algorithms of parallel random number generators[J]. Journal of Numerical Methods and Computer Applications, 2001(4):311-320
- [7] Lewis T G, Payne W H. Generalized Feedback Shift Register Pseudorandom Number Algorithm[J]. Journal of the ACM, 1973, 20(3):457-460
- [8] Ripley B D. Thoughts on pseudorandom number generators[J]. Journal of Computational and Applied Mathematics, 1990, 31(1):156-157
- [9] Wang En-dong, Zhang Qing, Shen Bo, et al. High-Performance Computing on the Intel Xeon Phi-How to Fully Exploit MIC Architectures[M]. China Water Power Press, 2012
- [10] L'Ecuyer P, Simard R. TestU01: A C Library for Empirical Testing of Random Number Generators[J]. ACM Transactions on Mathematical Software, 2007, 33(4):22