

文章编号:1001-9081(2006)12-2841-02

## 基于高级视觉特性的感兴趣区域判定算法

廖 斌,许 刚,刘春颖

(华北电力大学 电气与电子工程学院,北京 102206)

(nathan@ncepu.edu.cn)

**摘 要:**提出了一种快速确定视频场景中感兴趣区域的算法,其基本思想是通过对人类视觉系统 HVS 的某些高级视觉特性进行分析,对图像分段后获得的区域做出相应的视觉重要性估计,从而确定每个区域在主观意义上的视觉感兴趣值,该值的大小反映了其重要性程度。实验结果表明,利用该算法所获得的感兴趣区域基本符合人眼的主观判断,而且算法的复杂度较低,具有较高的实用价值。

**关键词:**感兴趣区域;图像分段;高级视觉特性

**中图分类号:** TP391.41 **文献标识码:** A

## ROI decision algorithm based on advanced visual properties

LIAO Bin, XU Gang, LIU Chun-ying

(School of Electrical Engineering, North China Electric Power University, Beijing 102206, China)

**Abstract:** A rapid algorithm to evaluate the importance of different regions in the video scenes was proposed. The principle of the algorithm is to make corresponding visual importance estimates of segmented regions by analyzing the advanced visual properties. As a result, the subjective visual ROI values were ascertained, which reflected the importance of each region. The experimental results show that the ROI values produced by the proposed algorithm accord with human subjective conclusions well. Besides, the computation cost of the algorithm is not high, and it has high utility value.

**Key words:** Region of Interest (ROI); segmented images; advanced visual properties

### 0 引言

在目前的图像质量评价体系当中,信噪比 PSNR 或均方误差 MSE 作为一种量化的图像质量评价指标得到广泛的采纳,但有时候它并不能够完全反映真实的人类视觉特性,因此随之而来出现了许多基于视觉特性的视频图像改进编码方案<sup>[1,2]</sup>。早期的视觉模型主要是利用人眼视觉对不同频带信号的感知差异,它们所关心的是压缩编码系统产生的误差是否为人眼可见(不考虑这些误差的分布位置),这些方法在一定程度上获得成功。然而现今许多应用系统当中,如窄带视讯会议系统、医疗以及遥感等,都允许图像的质量出现不均匀的分配,即允许某些人们感兴趣的区域的质量要相对高于其他区域。换句话说,人类视觉系统在观察图像的时候,对于图像中某些部分的误差关注程度明显是要高于其他周围区域的。那么,如何去识别出不同图像区域对人类视觉的重要程度则是我们研究的重点。

人类视觉注意力和眼球运动方面的研究<sup>[3,4]</sup>表明:对于一幅图像,即使给予观察者很长时间,人们通常也只会对其其中少部分区域表现出较强的兴趣。而且对于连续的运动图像而言,人眼所注意的对象一般



图 1 Plane 512 × 512 × 8 bpp

具有高度的相关性。如图 1 所示的一幅自然图像当中,人眼首先关注的物体当然是飞机,而不是山峦,这主要包括两点原因:一是飞机位于图像的中央部分;二是飞机与周围的背景具

有较强的对比度。因此,如果图中的飞机出现失真很容易就让人发觉,人眼在它上面停留的时间最长;相反,如果失真出现在山峦部分,则不易察觉。这些人眼对图像内容的认知因素都是以前的视觉模型没有考虑到的。

### 1 视觉关注模型

一般人都认为人脑具备很强的并行处理能力,然而它在处理复杂的视觉场景分析当中,串行处理过程却是必不可少的。由于人类视觉神经元所接受到的信息量都是按照  $10^8$  bit/s 为单位计算的,它远远超过了人脑可能的处理能力。为了有效地处理周围环境中所出现的大量信息,人类视觉系统 HVS 在观察场景时选择了不同的分辨率来区别对待这些信息。虽然人眼的视野范围很宽,水平范围能达到  $180^\circ$ ,垂直范围到达  $140^\circ$ ,但是视觉最为敏锐的部分却只有很窄的一部分(位于黄斑区范围内),大约只有  $2^\circ$ 。为了准确地辨别环境内的不同对象,人们必须控制眼球的转动,让视觉注意力停留在感兴趣的区域。为了确定感兴趣的区域,人眼需要进行快速扫描活动对场景进行粗略的浏览,人眼的快速扫描运动大约每  $100\text{ms} \sim 500\text{ms}$  就要发生一次,而人们的视觉关注机制将使得这些扫描活动具有明显的针对性。

### 2 影响人类视觉注意的因素

一般来讲,人类的注意力以及眼球的运动过程是同时受到高层次和低层次的因素控制的。高层次因素一般包含了来自记忆的反馈过程以及类似模式匹配的活动。低层次处理过程速度更快,就如同条件反射一样,可以说是一种自然属性。

收稿日期:2006-06-26;修订日期:2006-08-27

作者简介:廖斌(1974-),男,湖北武汉人,讲师,博士,主要研究方向:数据压缩技术,视频图像编码;许刚(1963-),男,安徽合肥人,教授,博士,主要研究方向:信号处理,语音及视频编码;刘春颖(1973-),女,河北石家庄人,讲师,硕士,主要研究方向:人工智能,动态系统建模。

目前已经确认的影响视觉注意力的低层次因素主要包括:

**对比度:**区域对比度是一种很强的吸引视觉注意的因素,与相邻区域具有很高对比度的区域往往显得更为重要。

**尺寸:**大面积区域比小区域更容易吸引视觉注意。当然,存在一个饱和度,一旦区域面积超过它,其重要性就会明显降低。

**形状:**狭长形状的区域同具备相同对比度和面积的区域相比更容易引起人们的视觉注意。

**运动:**当人们观察运动图像序列的时候,通常会专注于某些运动对象,想要搞清楚运动的细节。

除此之外,还有其他的一些低层次因素包括方向、颜色和亮度等。而已经发现的影响视觉注意力的高层次因素主要与人类的主观认知因素有关,它包括:

**位置:**人们在观察图像时,通常首先都会把注意力定格在图像中央的 25% 的区域范围之内。

**前景/背景:**人们的视觉对前景和背景的区分通常比较敏感,都会把注意力停留在前景对象上,而忽略背景的具体内容。

**人物:**人们对场景中出现的物形往往给予更多的关注,尤其是它的眼睛、脸部、嘴巴和手等肢体。

### 3 感兴趣区域自动判定算法

虽然影响人们视觉注意力的因素有很多,但是缺乏足够的数据来确定这些因素之间的加权关系。这意味着,即使某些因素明显具有更高的重要性,我们也无法衡量它相对于其他因素而言究竟有多重要。正是由于缺乏足够的信息,使得我们在分析视觉重要性时不得不考虑多种因素。由于无法准确度量这些因素之间的相对重要性,本文假设这些因素之间是相互独立的。

下面给出图像的感兴趣区域自动判定算法的主要计算步骤:

1) 将输入的原始灰度图像进行图像分割,得到若干图像区域  $Region(i)$ ,  $0 < i < N$ ,  $N$  为图像区域个数。

2) 计算每个图像区域针对不同因素的重要性,计算公式分别为:

(a) 对于任何一个区域  $R_i$ ,其与相邻区域的对比度因子的计算公式如下:

$$W_1(R_i) = \bar{Y}(R_i) - \bar{Y}(R_{i-NE}) \quad (1)$$

其中  $\bar{Y}(R_i)$  表示区域  $R_i$  的平均亮度值,而  $\bar{Y}(R_{i-NE})$  则表示所有相邻区域的平均亮度值。最后将对对比度因子归一化到  $[0,1]$  范围内。

(b) 区域尺寸因子计算公式如下:

$$W_2(R_i) = \max\left(\frac{Area(R_i)}{AR}, 1\right) \quad (2)$$

其中  $AR$  是一个常量,其值为图像总面积的 1%,目的是为了实现在区域面积的饱和度,防止给予面积过大的区域的权值过高。

(c) 区域形状因子计算公式如下:

$$W_3(R_i) = \frac{[Border(R_i)]^p}{Area(R_i)} \quad (3)$$

其中  $Border(R_i)$  表示区域  $R_i$  与其他区域交界的像素个

数,即区域边界的长度。而常量  $p$  的取值为 1.75,这样可以有效地区分区域的形状,使得狭长区域的重要性相对高于圆形区域。

(d) 区域位置因子计算公式如下:

$$W_4(R_i) = \frac{Center(R_i)}{Area(R_i)} \quad (4)$$

其中  $Center(R_i)$  表示区域  $R_i$  中处于图像中央 25% 范围之内的像素个数。这样,如果区域的所有像素都处于中央位置,它的权值就会是 1。而完全处于中央范围以外的区域的权值将会置为 0。

(e) 图像中背景区域的检测办法是通过分析区域中有多少比例的像素分布于图像的边界之上。前景/背景因子计算公式如下:

$$W_5(R_i) = 1 - \min\left(\frac{RBorder(R_i)}{ImageBorder(R_i)}, 1\right) \quad (5)$$

其中  $RBorder(R_i)$  表示区域  $R_i$  与图像边界接壤的像素个数,而  $ImageBorder(R_i)$  则表示图像边界的像素数目,它等于  $2 \times (width + height) - 4$  ( $width, height$  分别表示图像的宽度和高度)。

3) 将上述几种因子进行综合,得到最终的视觉重要评价。由于这些因素之间的相对重要性不可知,因此采用等加权方式。另外,为了使得某些具有较高值的视觉因子能够对最终的结果产生更重要的影响,我们采用平方求和的方式,而不是平均求和:

$$W(R_i) = \sum_{k=1}^5 W_k^2 \quad (6)$$

最后将最终获得的评价结果归一化到  $[0,1]$  范围内。

需要指出的是:在步骤 1 当中,采用的图像分割算法是 Split-Merge 算法<sup>[6]</sup>,该算法计算效率较高且分段结果也能够满足需求。Split-Merge 分割算法主要利用区域的方差作为区域分裂和合并的准绳,根据大多数图像分割的结果判断,将方差的门限值设为 200 是较为合理的。从实际情况来看,该分割算法的分割结果中通常会出现很多的尺寸很小的区域,为了避免过度分割,我们将尺寸小于 16 个像素的区域与平均亮度值最相近的相邻区域合并,这样可以减少分割区域的数量。

### 4 实验结果

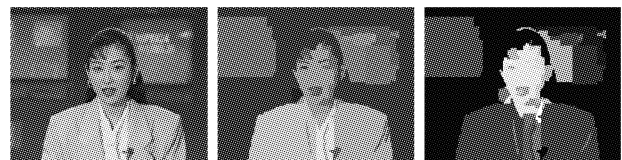


图2 News 场景判别结果



图3 Carphone 场景判别结果

利用本文给出的图像感兴趣区域自动判定算法对一些标准的国际视频序列进行了测试,如 News 序列和 Carphone 序列。图2给出了对视频序列 News 的测试结果,其中图2(b)

(下转第 2847 页)

绘制历史信息的应用,如利用多帧(本文使用了一帧)的历史来进行下一帧的预测,以提高算法有效性;甚至保留所有绘制历史,以某种智能型的决策达到负载平衡。二是将本文算法与文献[4]的方法进行混合,有选择地使用几何数据,以达到有效性和开销两利的效果。这些工作将在未来进一步展开。

#### 参考文献:

- [1] DAVIS T, CHALMERS A, JENSEN HW. Practical Parallel Processing for Realistic Rendering[ A]. ACM SIGGRAPH[ C]. 2000.
- [2] MOLNAR S, COX M, ELLSWORTH D, *et al.* A Sorting Classification of Parallel Rendering[ A]. IEEE computer graphics and applications[ C]. 1994. 23 - 31.
- [3] HUMPHREYS G, ELDRIDGE M. WireGL: A Scalable Graphics System for Clusters[ A]. Proceedings of ACM SIGGRAPH[ C]. 2001.
- [4] SAMANTA R, ZHENG J, FUNKHOUSER T, *et al.* Load Balancing for Multi-Projector Rendering Systems[ A]. Proceedings of the SIGGRAPH/Eurographics Workshop on Graphics Hardware[ C]. 1999.
- [5] MONTRYM JS, BAUM DR, DIGNAM DL, *et al.* InfiniteReality: A Real-Time Graphics System[ A]. Proceedings of SIGGRAPH '97[ C]. 1997. 293 - 302.
- [6] CROCKETT TW. Design Considerations for Parallel Graphics Libraries[ R]. ICASE Report No. 94 - 49, 1994.
- [7] MOLNAR S, EYLES J, POULTON J. PixelFlow: High-Speed Rendering Using Image Composition[ A]. Proceedings of SIGGRAPH '92[ C]. Chicago, Illinois, 1992. 231 - 240.
- [8] MITRA T, CHIUH T. Implementation and Evaluation of Parallel Mesa Library[ A]. IEEE International Conference on Parallel and Distributed Systems[ C]. 1998.
- [9] MOLL L, SHAND M, HEIRICH A. Sepia: Scalable 3D Compositing Using PCI Pallette[ A]. Laurent Proceedings of the Seventh Annual IEEE Symposium on Field-Programmable Custom Computing Machines[ C]. 1998.
- [10] SHI JY, JIN ZF. A Survey On Parallel Polygon Rendering[ J]. Computer Aided Design and Computer Graphics, 2003, 15(6): 637 - 642.
- [11] YANG J, SHI JY, JIN ZF, *et al.* Design and Implementation of A Large-scale Hybrid Distributed Graphics System[ A]. Eurographics Workshop on Parallel Graphics and Visualization[ C]. Saarbruecken, Germany, 2002.
- [12] ELDRIDGE M, IGEHY H, HANRAHAN P. Pomegranate: A Fully Scalable Graphics Architecture[ A]. Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH[ C]. New Orleans, Louisiana, 2000. 443 - 454.
- [13] MUELLER C. The Sort-First Rendering Architecture for High-Performance Graphics[ A]. Proceedings of the 1995 Symposium on Interactive 3D Graphics[ C]. 1995. 75 - 82.
- [14] DOUGLAS R. A Load Balanced Parallel Scanline Z Buffer Algorithm for the iPSC Hypercube[ A]. Proceedings of Pixim 88[ C]. Paris, France, 1988. 177 - 192.
- [15] DANIEL W. Animac: A Multiprocessor Architecture for Real-Time Computer Animation[ D]. Ph. D. dissertation, California Institute of Technology, 1985.
- [16] WHITMAN S. Dynamic Load Balancing for Parallel Polygon Rendering[ J]. IEEE Computer Graphics and Applications, 1994, 14 (4).

(上接第 2842 页)

给出了经过区域合并处理后的分段图像,图 2(c)给出了最终分析得到的视觉重要性评价,该图像中的区域的亮度值反映其在视觉上感兴趣的程度,亮度值越高则感兴趣程度越高。它基本上能够反映出人眼视觉的主观特性,例如播音员的脸

部和衣襟部分亮度值较大,说明其视觉敏感度较高,这与人类的视觉特性基本吻合。同时,表 1 列出了 News 场景部分区域的视觉感兴趣值及相关特性如区域尺寸和平均亮度。对于较复杂的场景如视频序列 Carphone,该算法也能给出较为理想的结果,如图 3 所示。

表 1 实验结果的部分数据统计

区域编号	1	2	3	4	5	6	7	8	9	10	11	12	...
面积	26	29	31	23	136	1099	35	66	49	32	41	57	...
亮度均值	122	109	93	111	105	135	108	125	92	85	95	102	...
视觉评价	1	1	0.95	0.93	0.91	0.90	0.87	0.87	0.85	0.84	0.82	0.81	...

## 5 结语

通过对人眼在观察视频场景的一些低级视觉特性和高级视觉特性的综合分析,提出了一种基于高级视觉特性的感兴趣区域自动判定算法,它结合了不同的视觉因素,能够自动地计算每个区域的视觉重要性估计值,这个估计值反映了每个区域的视觉感兴趣程度。最后的实验结果表明:该算法所获得的结果基本上符合人眼主观视觉特性,且该算法实现的计算复杂度不高,因此具有较强的实用价值。同时,该算法也存在一些不足,如对细节较多的图像判别效果不是很理想,如何结合更好的图像分割算法,同时加入人脸检测的方法,使得感兴趣区域判定更为准确,是值得进一步研究的问题。

#### 参考文献:

- [1] 胡栋,郑宝玉. 数字视频关键技术的若干新进展[ J]. 通信学报, 2003, 24(7).
- [2] 一种新的基于率失真优化的感兴趣区域编码[ J]. 计算机应用, 2005, 25(1).
- [3] TRAN TD, SAFRANEK R. A locally adaptive perceptual masking threshold model for image coding[ A]. ICASSP[ C]. Atlanta, 1996.
- [4] WOLF S, WEBSTER A. Subjective and objective measures of scene criticality[ A]. ITU meeting on subjective and objective audiovisual quality assessment methods[ C]. Turin, 1997.
- [5] VILLEGAS P, MARICHAL X, SALCEDO A. Objective evaluation of segmentation masks in video sequences[ A]. WIAMIS'99[ C]. Germany, 1999. 85 - 88.
- [6] HOROWITZ SL, PAVLIDIS T. Picture segmentation by a directed split-and-merge procedure[ R]. Technical report, Department of Electrical Engineering, Princeton University, N. J. 08540, 1975.