基于视觉注意力模型的显著性提取

张 杰，魏 维
（成都信息工程学院 计算机系，四川 成都 610225）

摘 要：对现有基于注意力机制的静态显著计算和动态显著计算技术进行综述。它主要包括两部分：静态图像的显著性提取和动态图像的显著性提取。静态显著计算首先介绍了Itti和Stentiford静态显著性提取模型，然后分析了基础分割的注意力模型技术。动态显著性提取中的两个动静结合的注意力模型、强注意力偏向融合和基于运动优先的注意力模型。介绍了一些视觉注意力模型，并对其进行了讨论。探讨了各种模型的优缺点及应用。为视觉注意力模型在图像检索、人机交互、视频监控等领域提供了一定的基础。

关键词：视觉注意力；显著性提取；静态显著；动态显著
中图分类号：TP391 **文献标识码**：A **文章编号**：1673－629X(2010)11－0109－05

Saliency Extraction Based on Visual Attention Model

ZHANG Jie, WEI Wei
(Dept. of Computer, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Overview of the paper is proposed of static and dynamic saliency calculation based on attention mechanism. It mainly consists of two parts: saliency extraction of static image and saliency extraction of dynamic image. Static saliency calculation firstly introduces visual attention models of Itti model and Stentiford model, then analyze the technology of attention model based on segments. The model of saliency extraction in dynamic image is the fusion of static attention model and dynamic attention model, strong attention bias fusion and attention model based on motion. The paper introduces some visual attention models. Have a certain discussion in these visual attention models, and point out their advantages and disadvantages and the application. It provides some basis in image retrieval, human computer interaction, video surveillance and other fields for the visual attention model.

Key words: visual attention; saliency extraction; static saliency image; dynamic saliency image

0 引 言

视觉注意力模型是一种用计算机来模拟人类视觉注意力系统的模型，在一幅图像中提取人眼所能观察到的引人注意的焦点，相对于计算机而言，就是该图像的显著性区域。近年来，视觉注意力模型的研究已成为焦点，它可以应用到基于感兴趣区域的图像检索、计算机视觉、关键帧的提取、图像压缩等领域。

人类的视觉系统拥有图像理解、识别、处理的能力，让计算机模拟视觉系统建立视觉注意力模型是图形图像处理领域中的研究热点[1]。根据生物视觉理论，视觉注意机制（visual attention mechanism）人眼对信息的处理不是均衡的，它会自动地对感兴趣区域进行处理，提取出有用的信息，不感兴趣区域则不作处理[2]。视觉注意机制就是使人们能够在复杂的视觉环境中快速定位感兴趣目标。

建立视觉注意力模型（visual attention model, VAM）是让计算机模拟人类的视觉注意机制得到图像中最容易引起人们注意的部分[3]。研究表明，人们比较关注感兴趣的区域。检测感兴趣区域就是利用视觉注意力模型得到图像中显著度较高的区域[4]。因此，文中提出了几种提取显著性区域的视觉注意力模型，并对此进行了分析与探讨。

收稿日期：2010－03－29；**修回日期**：2010－06－09
基金项目：成都信息工程学院引进人才启动科研项目（KYTZ200914）；成都信息工程学院发展基金项目（CSRF200803）；四川省教育厅青年基金项目（2006B063）
作者简介：张 杰（1984－），女，硕士生，研究方向为图像处理、媒体显著性研究等；魏 维，副教授，研究方向为图像处理、语义分析等。

1 视觉注意力

1.1 提取静态图像显著性的视觉注意力模型

1.1.1 Itti注意力模型

视觉心理学研究表明，人类视觉系统选择性注意机制主要包括：

（1）采用自底向上（bottom－up）控制策略的预注

意机制，属于低级的认知过程。它是基于视觉注意力模型来计算图像的显著性，不考虑特定的认知任务对提取图像显著性的影响。

(2)采用自顶向下(top－down)控制策略的注意机制，属于高级的认知过程。根据任务需求提取图像的显著性，对其进行有意识的处理，从而得到人们想要的感兴趣区域。

Itti 注意力模型属于自底向上控制策略的预注意机制，它是由 Itti 等人[5~7]提出的，是比较经典的视觉注意力模型之一。该模型的基本思想(如图 1 所示)是，在图像中通过线性滤波提取颜色特征、亮度特征和方向特征，通过高斯金字塔、中央周边操作算子(center－surround differences)和归一化处理后，形成 12 张颜色特征地图、6 张亮度特征地图和 24 张方向特征地图。将这些特征地图结合和归一化处理后，分别形成颜色、亮度、方向关注图(conspicuity maps)，三个特征的关注图线性融合生成显著图(saliency map)通过两层的赢者取全神经网络(winner－take－all，WTA)，得到显著区域，最后通过返回抑制机制，抑制当前显著区域，转而寻找下一个显著区域。

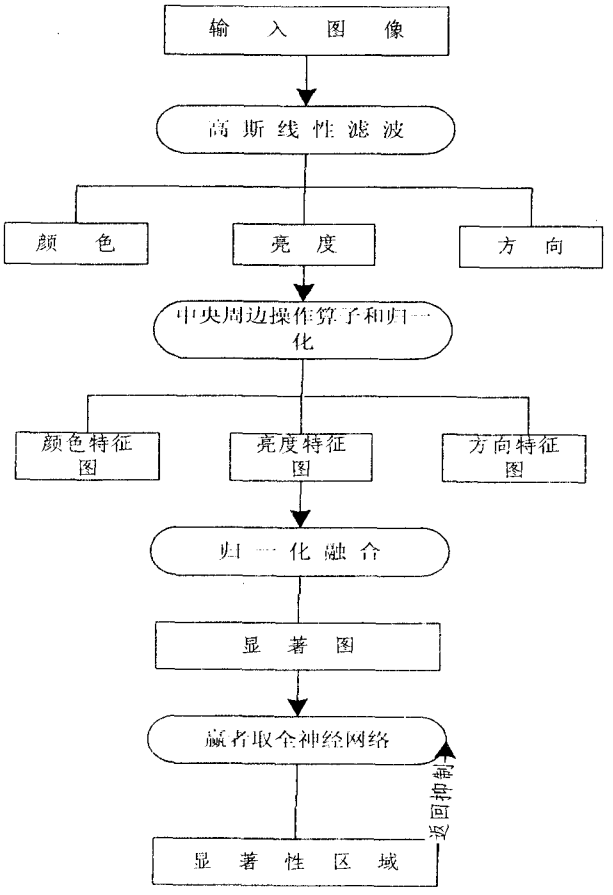


图 1 Itti 模型框架

当输入一张自然风景图片时，该模型第一次执行的结果如图 2 所示。

从图 2 可以看到，当执行一次程序之后其感兴趣区的大小是在注意力焦点处指定一定形状的区域(一般用圆，如原图中的亭子上部的黄色圆圈)来表示的，这与人眼所观察到的实际目标相比有一定的偏差。但 Itti 模型在静态图片中提取显著性区域是比较经典的，以后的许多模型都是在此基础上发展起来的。

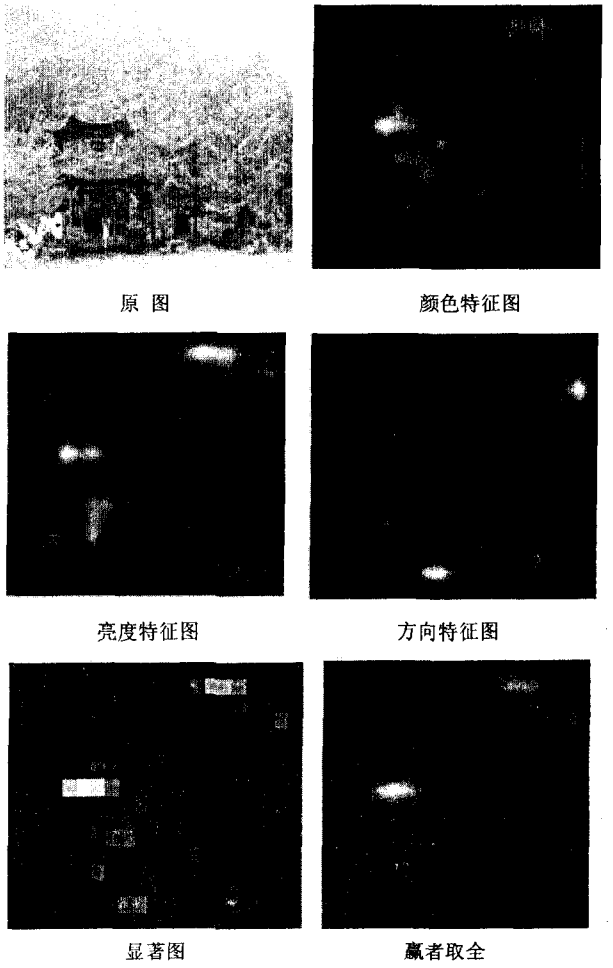


图 2 Itti 模型执行结果

1.1.2 Stentiford 注意力模型

Stentiford 等人[8,9]在图像的检索中应用到了视觉注意力模型，图像中的显著区域用视觉注意力图(visual attention map，VA)来表示。若图像中某像素和它周围区域的特征(如形状、颜色等)在图像其它相同形态区域中出现的频率越多，该像素的 VA 值越低，反之 VA 值越高[10]。其 VA 图如图 3 所示，类似于经典的 Itti 模型中的显著图。

1.1.3 以分割为基础的注意力模型

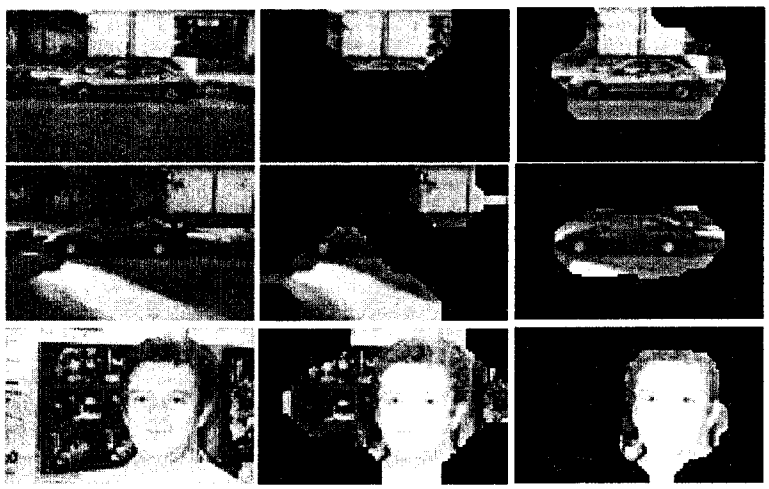
提取显著对象的大多数算法是建立在像素基础之上的。大部分的算法是很有效的，然而它们存在两个缺点。首先，运算代价太高。为了提取显著对象所呈现的特征，必须用多尺度特征融合策略[11]或复合矩形

扫描策略[12]。这些处理过程都是很消耗时间的。其次,以特征为基础的像素一般都不能反映整体的显著对象,它们很可能在杂乱无章的背景下检测不到显著对象。

(a)原图　　(b) VA 图

图 3　Stentiford 模型

为了解决以上两个问题提出了以分割为基础的新颖的显著对象检测算法,将其命名为 SODS(salient object detection based on segments)[13]。首先用有效分割的方法,分割一个图像。然后在分割的基础上提取多尺度对比差、中央－周边直方图和颜色空间分布特征,将这三种特征地图进行线性组合。试验证明,该算法比文献[12]中的要快,而且在杂乱无章的背景下也能检测到显著对象。如图 4 所示。



(a)原图　　(b) 刘铁模型　　(c)SODS 模型

图 4　提取显著对象

1.2　提取动态图像显著性的视觉注意力模型

1.2.1　基于大脑两条处理路径的动态场景视觉注意力模型

该模型基于在动态场景中大脑的两条处理路径、Ungerleide[14]指出视觉注意力是由大脑中的两条路径控制的,背部的路径处理运动刺激;腹部的路径处理颜色、亮度、方向和其它方面的刺激。Dobkins 和 Albright[15]揭示了这两条路径的相互作用。本模型的结构如图 5 所示[16],静态显著图是表示亮度、颜色和方向等静态特征。动态显著图是表示动态特征的,两个地图相互融合成刺激整合神经网络(integrate－and－fire neural network, IFNN)。

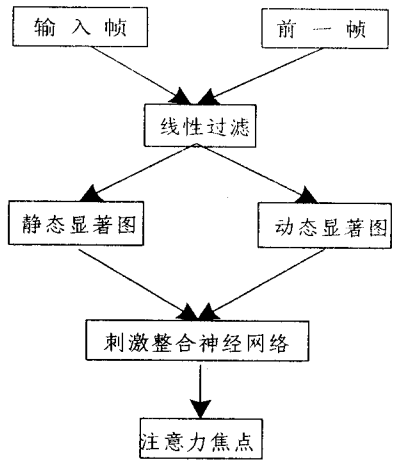


图 5　模型结构图

通过实验,将本模型与 Itti 与 Ouerhani 模型相比较,该模型更接近人类的注意力。

1.2.2　偏向强注意力的启发式融合策略

在动态图像中,最引人注意的一般是运动区域[17~19]。确定运动区域简单而有效的方法是计算两个连续帧之间方位的差异,如果某点的位置差比开始时大,则说明该点是个运动像素点。一个区域中包含的运动像素点越多,也就是这些运动像素点的位置差异越大,它的运动速度可能越大,因此,很容易被人们觉察到。但有时候在一幅动态图像中,引人注意的不是运动对象,而是在中间位置或色彩对比度很大的静态对象,这就需要引入偏向强注意力的启发式融合策略。

一个强烈的注意力提示常常能够引起人们的兴趣,而并非所有的注意力提示或所有注意力的平均值。鉴于此,设计了一种偏向强注意力的启发式融合策略("strong attention bias fusion", SABF)[20]。SABF 可用下式表示:

$$A^k(R_m^k) = E(A) + \delta \cdot \max(A_1, A_2) \cdot \delta =$$

$$\frac{1}{2}(A_1 + A_2) + \frac{1}{2}\delta \cdot \max(A_1, A_2) \cdot |A_1 - A_2|$$

AF("average fusion",平均值融合法,取静态注意力和动态注意力二者的平均值)和 SABF 两个融合策略的不同在与"偏向"的部分[21]。例如:如果一个对象的静态注意力是 1,而运动注意力是 0,通过 AF 可知其整体的注意力值是 0.5。因此就不能容易的决定这个对象是否有吸引力。然而,通过 SABF,它的整体注意力值可以达到 0.95,它就属于感兴趣对象。SABF

既涉及平均因素又涉及偏向强注意力因素，因而更具合理性。

1.2.3 基于运动优先的视觉注意力模型

在提取了运动和静态显著区域后，将运动和静态显著区域合成一个视觉注意度，如下式：

$$A = \omega_s \cdot A_s + \omega_T \cdot A_T$$

其中，ω_s，ω_T 分别为运动和空间显著度权重，研究文献[22]、文献[23]表明，采用固定权重值无法自适应根据视频内容变化调整运动和空间显著度比例。因此提出了一种运动优先的非线性混合模式将静态和运动显著度进行动态混合。当运动加强时，其运动权重迅速加大，静态图像权重迅速减小，但当运动强度达到一定强度时，其运动权重增加应该减缓。则运动优先原则定义为下式：

$$\omega_T = \bar{A}_T \cdot \exp(1 - \bar{A}_T)$$

其中，$\bar{A}_T = \max(A_T) - \mathrm{mean}(A_T)$，即 $\bar{A}_T$ 为 A_T 最大值减去 A_T 平均值，能够自适应根据运动的反差变化调节 $\bar{A}_T$。ω_T，ω_s 随 $\bar{A}_T$ 变化的曲线如图6所示。由图6可知，随着运动显著度的加强，运动权重值快速增加，同时静态权重值快速减少，当运动显著度达到一个值后，为了兼顾静态显著图，其 ω_T 增长速度减缓，而 ω_s 减少程度也减缓，这样的设计既考虑到运动优先，又同时兼顾静态部分。

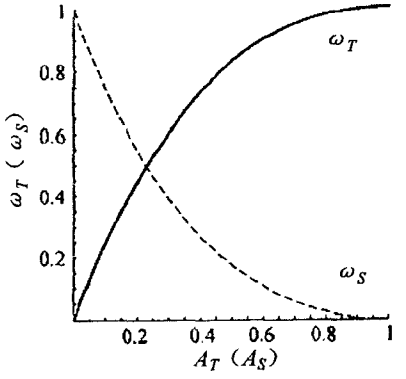


图6 运动优先的动态权值变化图

2 讨 论

Itti模型：该模型首先融合图像的一些低层视觉特征生成显著图(saliency map)，然后使用一个动态的神经网络按照显著度递减的顺序依次发现图像中的注意点。在静态图片提取显著性，即在一张图片中寻找感兴趣区域时，该模型是比较经典的。以后许多模型都是建立在此基础上的，如：桑农[24]将这一模型应用到可见光图像中的RGB颜色空间的目标检测中，叶聪颖等人[25]将这一模型转换到HIS颜色空间，并应用到海上目标检测中，都取得了较好的效果。但在实际的显著性提取中，Itti模型检测到的显著性区域跟实际人眼看到的显著性区域有一定的差距，该模型一般是用圆圈标注的。于志文等人[26]提出的基于实时聚类的视觉注意力区域的提取，提取出的显著性区域比较接近客观事物的形状，该模型填补了传统的视觉注意力区域和高层语义之间的鸿沟。

Stentiford模型：能很好地识别出显著特征和非显著特征，但是如果图像区域不够显著，可能造成不理想的结果。

以分割为基础的注意力模型：将分割应用到显著对象检测上不是首次尝试，但是大多数运用分割的方法是作为以像素为基础的方法的增强或补充。例如：F. Liu et al[11]用Itti模型得到显著地图和区域信息来消除令人误解的线条。Vidya Setlur模型也是建立在Itti模型基础之上的。在文献[11]中作者确实以检测显著对象的方法为基础，介绍了一个较为完美的区域，但是它太简单而且不能满足操作的需求。该模型达到了更好的操作需求，运行时间短，抗干扰能力强。不过，从图4可以看出，该模型提取的显著对象仍不能很好地满足人眼所达到的提取程度。

基于大脑两条处理路径的动态场景视觉注意力模型：在动态场景中提取显著性区域，基本都是采用动态特征和静态特征相结合的方法。基于大脑两条处理路径的动态场景视觉注意力模型是通过将动态显著图和静态显著图相结合成刺激整合神经网络来提取显著区域的，该模型能动态地定位空间中的注意力焦点，它的处理过程类似于自然的注意力。如果能将影响注意力的对象尺寸和不同运动对象的竞争考虑进去，该模型将有进一步的提高。

偏向强注意力的启发式融合策略：该模型是采用静态注意力和动态注意力相融合的方法。它解决了平均值融合法的缺陷，当静态注意力和动态注意力的值都大时，平均值融合法所提取的显著性区域跟人类的注意力相符。但当二者之一的注意力值很大，另一个较小时，用平均值融合法就检测不到显著区域，而此时人类的注意力则是偏向注意力值大的那一方，此时就需要用偏向强注意力的启发式融合策略。

基于运动优先的视觉注意力模型：该模型分别提取空间和运动显著度，并采用运动优先原则将时空显著度进行动态混合形成视觉注意度[27]。试验证明，该方法提取的显著区域符合人类的视觉注意力系统。

3 结束语

介绍了提取显著性的视觉注意力模型，通过这些模型，既可以提取静态图像的显著性，又可以提取动态

图像的显著性。目前,静态显著性的研究已经比较成熟,如:典型的 Itti 模型。需要对动态显著性做进一步的研究和发展。利用视觉注意力模型提取显著性区域,为进一步实现基于人类感兴趣区域的图像检索奠定了基础。研究视觉注意力模型对图像压缩、人机交互和视频监控等技术的发展具有重要意义。

参考文献:

[1] Bulthoff H H, Lee S W, Poggio T, et al. Biologically Motivated Computer Vision[M]. New York: Springer Publishing Company, 2003: 150 - 159.

[2] Huang L, Pashler H. Working memory and the guidance of visual attention: Consonance - driven orienting[J]. Psychonomic Bulletin & Review, 2007, 14(1): 148 - 153.

[3] 张　菁,沈兰荪,高静静.基于视觉注意模型和进化规划的感兴趣区域检测方法[J].电子与信息学报,2009,31(7):1646 - 1652.

[4] 张　菁, 沈兰荪, Feng D D. 基于视觉感知的图像检索的研究[J]. 电子学报, 2008, 36(3): 494 - 499.

[5] Itti L, Koch C, Niebur E A. Model of saliency—based visual attention for rapid scene analysis[J]. IEEE Trans on PAMI, 1998, 20(11): 1254 - 1259.

[6] Itti L, Koch C. Computational Modeling of Visual Attention [J]. Nature Reviews Neuroscience, 2001, 2(3):194 - 203.

[7] Navalpakkam V, Itti L. Modeling the Influence of Task on Attention[J]. Vision Research, 2005, 45(2):205 - 231.

[8] Stentiford F W M. An Attention Based Similarity Measure with Application to Content Based Information Retrieval[C] // In Proceedings of the Storage and Retrieval for Media Databases conference, SPIE Electronic Imaging. Santa Clara, CA:[s. n.], 2003.

[9] Bamidele A, Stentiford F W M. An Attention Based Similarity Measure Used to Identify Image Clusters[C]// In Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantics & Digital Media Technology. London:[s. n.], 2005.

[10] 高静静,张　菁,卓　力,等.应用于图像检索的视觉注意力模型的研究[J].测控技术,2008,27(5):19 - 21.

[11] Liu F, Gleicher M. Region enhanced scale - invariant saliency detection[C] // In Proceedings of IEEE ICME. [s. l.]:[s. n.], 2006.

[12] Liu Tie, Sun Jian. Learning to detect a salient object[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2007). [s. l.]:[s. n.], 2007:1 - 8.

[13] ZHUANG Liansheng, TANG Ketan, YU Nenghai, et al. Fast Salient Object Detection Based on Segments[C]// International Conference on Measuring Technology and Mechatronics Automation of IEEE. [s. l.]:[s. n.],2009:469 - 472.

[14] Ungerleider L G, Haxby J V. 'What' and 'where' in the human brain[J]. Current Opinion in Neurobiology, 1994(4): 157 - 165.

[15] Dobkins K R, Albright T D. What happens if changes color when it moves: The nature of chromatic inprt to macapre misual Area MT[J]. Journal of Neuroscience, 1994, 14 (8): 4854 - 4870.

[16] Chen Jiawei, Lin Kunhui, Zhou Changle, et al. A Visual Attention Model for Dynamic Scenes Based on Two - Pathway Processing in Brain[C]// Fourth International Conference on Natural Computation. [s. l.]: IEEE Computer society, 2008: 128 - 132.

[17] Kim M, Choi J G, Kim D, et al. A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio - temporal information[J]. IEEE Trans. Circuits Syst. Video Technol., 1999,9(8):1216 - 1226.

[18] Tsaig Y, Averbuch A. Automatic segmentation of moving objects in video sequences: a region labeling approach[J]. IEEE Trans. Circuits Syst. Video Technol., 2002, 12(7):597 - 612.

[19] Kim C, Hwang J - N. Fast and automatic video object segmentation and tracking for content - based applications[J]. IEEE Trans. Circuits Syst. Video Technol., 2002, 12(2): 122 - 129.

[20] Mezaris V, Kompatsiaris I, Strintzis M G. Video object segmentation using Bayes - based temporal tracking and Trajectory - based region merging[J]. IEEE Trans. Circuits Syst. Video Technol., 2004, 14(6):782 - 795.

[21] Han Junwei. Object Segmentation from Consumer Videos: A Unified Framework Based on Visual Attention[J]. IEEE Transactions on Consumer Electronics, 2009, 55(3): 1597 - 1605.

[22] Ma Y F, Hua X S. A generic framework of user attention model and its application in video summarization[J]. IEEE Transactions on Multimedia, 2005,10(7):907 - 919.

[23] Zhai Yun, Shah M. Visual attention detection in video sequences using spatiotemporal cues[C]// In: Proceedings of 14th Annual ACM International Conference on Multimedia. Santa Barbara, CA, USA:[s. n.], 2006: 815 - 824.

[24] 桑　农,李正龙,张天序.人类视觉注意机制在目标检测中的应用[J].红外与激光工程,2004, 33(1):38 - 41.

[25] 叶聪颖,李翠华.基于 HIS 的视觉注意力模型及其在船只检测中的应用[J].厦门大学学报:自然科学版,2005, 44(4):484 - 488.

[26] Yu Zhiwen, Wong Hau - San. A Rule Based Technique for Extraction of Visual Attention Regions Based on Real - Time Clustering[J]. IEEE Transactions on Multimedia, 2007, 9(4):766 - 784.

[27] 蒋　鹏,秦小麟.基于视觉注意模型的自适应视频关键帧提取[J].中国图像图形学报,2009, 14(8):1650 - 1655.