

## 目标跟踪的视觉注意计算模型

曾志宏<sup>1</sup>, 周昌乐<sup>1</sup>, 林坤辉<sup>2</sup>, 曲延云<sup>1</sup>, 陈嘉威<sup>1</sup>

(1. 厦门大学信息科学与技术学院, 厦门 361005; 2. 厦门大学软件学院, 厦门 361005)

**摘要:** 借鉴心理学中有关视觉注意的研究成果, 提出一种应用于视频图像序列的注意焦点计算模型。针对 Itti 模型算法计算复杂度高、运算速度慢等缺点, 通过对注意焦点的跟踪来降低计算量, 实现注意焦点的实时监测。利用相邻帧之间的相似性, 通过加权颜色直方图, 以最大匹配度作为预测模型, 得到下一帧注意焦点的位置。实验结果表明, 该注意焦点计算鲁棒性较好, 可以满足动态场景实时监测的需要。

**关键词:** 注意焦点; 跟踪; 相似性; 颜色直方图

## Visual Attention Computational Model Based on Tracking Target

ZENG Zhi-hong<sup>1</sup>, ZHOU Chang-le<sup>1</sup>, LIN Kun-hui<sup>2</sup>, QU Yan-yun<sup>1</sup>, CHEN Jia-wei<sup>1</sup>

(1. College of Information Science and Technology, Xiamen University, Xiamen 361005;

2. College of Software, Xiamen University, Xiamen 361005)

**【Abstract】** This paper proposes a model of the focus of attention in video sequences. The Itti model has the complex and slow computation, so it reduces computation to monitor the focus of attention through tracking the focus of attention. It uses the similarity between the adjacent frames, establishes the color histogram, selects the maximum similarity as predicable model, and gets position of the focus of attention in the next frame. The experimental results indicate that this model is effective in robustness and real-time.

**【Key words】** focus of attention; tracking; similarity; color histogram

## 1 概述

视觉注意机制作为视觉系统的一项重要心理调节机制, 是人类从外界输入的大量信息中选择特定感兴趣区域的一个关键技术, 由于可以在一定程度上实现有选择性地获取所关注目标的显著(saliency)信息, 从而大量降低信息处理量, 因此被广泛关注。选择性注意<sup>[1]</sup>计算模型是借鉴视觉注意机制提出和发展起来的。

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 2 个子过程:

(1) 采用自底向上(bottom-up)控制策略的注意机制, 该机制是基于输入景象的显著性计算的, 属于低级的认知过程。

(2) 采用自顶向下(top-down)控制策略的注意机制, 它通过调整选择准则, 以适应外界命令的要求, 从而达到将注意力集中于特定目标的目的, 属于高级的认知过程。

目前视觉心理学对(1)的研究较多, 采用(1)的视觉注意计算模型中以 Itti 等人提出的模型最具代表性<sup>[2-4]</sup>。该模型首先提取图像的多种初级视觉特征, 并通过 Center-Surround 算子计算各种特征在不同尺度下的视觉反差, 得到各自特征的显著图, 最后把不同特征对应的显著图进行融合, 从而得到反应视觉显著信息的注意焦点<sup>[5]</sup>。Itti 提出的模型在静态图像的显著区域检测上取得了良好的效果, 但是由于该模型计算复杂度较高, 运算速度较慢, 将该模型应用于动态场景之中很难满足实时性的要求。因此, 如何将视觉注意模型应用于动态场景之中, 实现快速准确的注意焦点计算, 为本文研究的主要内容。

## 2 模型的结构

本文提出了一种应用于视频图像序列的注意焦点计算模

型(如图 1 所示)。该模型由注意焦点计算、注意焦点跟踪、目标模型更新 3 部分组成。比较视频当前帧与前一帧之间的相似度, 如果两帧差异度大于阈值, 则采用 Itti 模型计算注意焦点, 并采用基于 HSI 的加权颜色直方图<sup>[6]</sup>生成目标模型; 如果两帧差异度小于阈值, 则跟踪由前一帧注意焦点生成的目标模型, 通过目标模型与候选模型之间的相似度匹配, 实现注意焦点的实时跟踪。

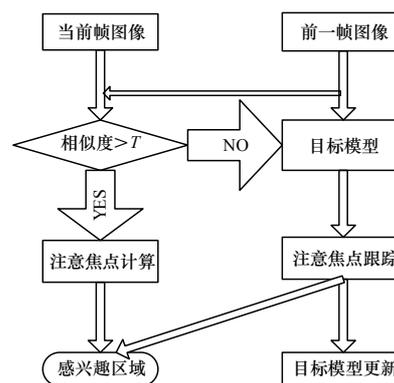


图 1 模型构架

## 2.1 相似度计算

通过定义一个帧运动量  $D_t$  来度量两帧之间的相似度:

**基金项目:** 国家“863”计划基金资助项目(2006AA 01Z129); 国家自然科学基金资助项目(60672018)

**作者简介:** 曾志宏(1982—), 男, 硕士研究生, 主研方向: 图像处理, 目标检测与跟踪; 周昌乐, 教授; 林坤辉, 副教授; 曲延云, 讲师; 陈嘉威, 博士

**收稿日期:** 2008-06-11 **E-mail:** xmuzzh82@hotmail.com

$$D_t = \frac{1}{c \times r} \sum_{y=1}^r \sum_{x=1}^c \frac{|I_2(x,y) - I_1(x,y)|}{255} \quad 0 \leq D_t \leq 1 \quad (1)$$

其中,  $I_1, I_2$  为两帧图像的灰度图;  $c \times r$  为视频帧的大小。

设定一个阈值  $T$ , 如果  $D_t > T$ , 说明两帧的差异较大, 则重新计算当前帧的注意焦点; 如果  $D_t < T$ , 则说明两帧相似度较高, 使用目标跟踪方法对上一帧的注意焦点进行跟踪, 作为结果输出。

## 2.2 注意焦点计算

利用 Itti 模型对输入的图像进行线性滤波, 提取亮度(intensity)、颜色(color)、方向(orientation)这些低层次的图像特征, 然后通过高斯金字塔和 Center-Surround 算子进行局部视觉反差的计算, 反差大的区域就代表容易吸引视觉注意力的区域, 此后再对不同尺度不同特征下的反差计算结果进行融合, 最终得到注意焦点, 如图 2 所示。



图 2 注意焦点

## 2.3 注意焦点跟踪

### 2.3.1 目标模型

现代研究表明, RGB 空间基于三色合成理论, 适合于彩色显像管等设备工作, 是不直观并且感知上不均匀的颜色空间, 光照很容易引起 RGB 值的变化; 为此选择 HSI 颜色空间, 3 个分量中  $H$ (Hue)代表色调,  $S$ (Saturation)代表颜色的饱和度,  $I$ (Intensity)代表亮度。该颜色空间以人眼对颜色的感知为基础, 具有较好的颜色感知一致性, 能方便地描述和控制彩色信息, 适合计算图像之间的颜色相似性, 另外具有平移不变性和缩放不变性等特征, 建立三维直方图。

颜色直方图描述了图像关于颜色的数量特征, 全图的颜色直方图算法比较简单, 有时可能 2 幅完全不同的图像具有完全一样的直方图, 不反映颜色位置信息, 为此引入加权颜色直方图, 即将图像分成几个子块, 对不同位置的像素赋予不同的权重, 位置与目标中心的距离越近, 其相应得到的权重越大, 增加了鲁棒性。假设注意焦点是中心为  $y_0$ , 半径为  $h$  的圆。

假设其中有  $n$  个像素用  $\{x_i\}_{i=1,2,\dots,n}$  表示, 特征值  $bin$  的个数为  $m$  个。则目标跟踪模型的加权颜色直方图即目标模型特征值  $u=1,2,\dots,m$  估计的概率密度为

$$q^{(n)}(y) = C \sum_{i=1}^n k\left(\left\|\frac{y_0 - x_i}{h}\right\|\right) \delta[b(x_i) - u] \quad (2)$$

其中,  $k(x)$  为核函数的轮廓函数, 本文选择 Epanechnikov 函数;  $h$  为核函数  $k$  的窗宽。由于遮挡或者背景的影响, 目标模型中心附近的像素比外物像素更可靠,  $k(x)$  对中心的像素给一个大的权值, 而远离中心的像素给一个小的权值。 $\delta(x)$  是 Kronecker Delta 函数, 其中  $b(x_i)$  为  $x_i$  处像素对应直方图中的颜色索引。  $C$  是标准化的常量系数, 使  $\sum_{i=1}^n q_u = 1$ , 因此:

$$C = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y_0 - x_i}{h}\right\|\right)} \quad (3)$$

### 2.3.2 候选模型

初始焦点的确定, 使得各帧中可能包含焦点的区域称为目标候选区域, 中心坐标为  $y$ , 也是核函数的中心坐标, 该区域中的像素用  $\{x_i\}_{i=1,2,\dots,n_h}$  表示, 参照式(2), 对候选区域的描述称为目标候选模型, 候选模型的特征  $u=1,2,\dots,m$  的概率密度为

$$p_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|\right) \delta[b(x_i) - u] \quad (4)$$

目标模型与候选模型之间的相似程度, 采用 Bhattacharyya 系数<sup>[7]</sup>作为相似性函数, 其定义为

$$\rho(p, q) = \sum_{i=1}^m \sqrt{p_u q_u} \quad (5)$$

其值在 0~1 之间,  $\rho(p, q)$  值越大, 表示 2 个模型越相似, 本文通过建立同心圆来对目标区域进行跟踪, 假设速度不会超过某个范围, 半径以 0~2 的放缩进行搜索, 在当前帧中的不同候选区域中计算得到候选模型, 选定  $\rho(p, q)$  最大候选区域, 即在本帧中目标的位置, 从而作为下一帧的目标位置, 并对目标模型更新。

## 2.4 目标模型更新

注意焦点在跟踪过程中可能会因为光照、变形等而变化, 因此对目标模型进行更新, 根据当前的结果按一定比例来对下一帧目标模型进行更新:

$$q_k = (1 - \alpha) q_{k-1} + \alpha p_{\max(\rho)} \quad (6)$$

其中,  $q_k$  是第  $k$  帧目标的颜色模型;  $p_{\max(\rho)}$  是匹配度 Bhattacharyya 系数最大的值, 也是匹配位置处候选目标的颜色模型;  $\alpha$  是设定的更新速度。

## 3 实验结果与分析

为了验证对 Itti 模型改进的有效性, 依照本文提出的算法对视频图像序列进行了测试。

本实验采用了 HIS 彩色空间建立的三维直方加权图, 对  $H, S, I$  进行了量化, 分成  $16 \times 16 \times 16$ , 首先采用了 Itti 模型计算出初始焦点, 然后利用相似性, 对注意焦点进行了有效的跟踪, 在真实的图像上进行了实验, 获得了较为满意的结果。

图 3 是本文算法与 Itti 算法的对比实验结果, 图 3(b)、图 3(e)是本文算法的检测结果, 图 3(c)、图 3(f)是 Itti 算法的检测结果。在检测效果上, 本文算法比 Itti 算法更加符合视觉感知: 在背景相对固定的情况下, 随着初始焦点的确定, 依据人的视觉注意机制, 对运动着的物体更加关注。在生成注意焦点时, Itti 算法在每一帧的多尺度特征提取中的耗时较多, 而本文算法由于充分利用了相邻帧之间的相似性, 无需在每一帧特征提取中耗时, 从而速度较快, 更加符合实时性。因此, 总的来说, 与现在常用的方法相比, 本文的方法具有以下优点:

(1)在整个注意焦点跟踪过程中, 受背景的影响不大, 适用于较为复杂的场景。

(2)利用了相邻帧之间的相似性, 减少计算复杂性, 速度较快。

(3)具有良好的鲁棒性, 可以解决焦点发生变形问题。从图 3(b)、图 3(e)中可以看出, 本方法对注意焦点的尺度、变形无关, 并且对光照等不敏感。

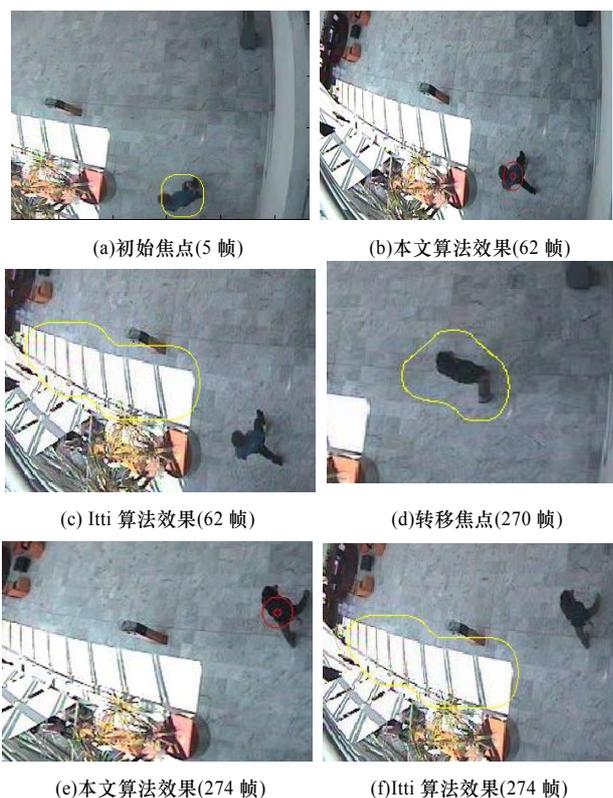


图3 本文算法效果

#### 4 结束语

本文在 Itti 等人提出的视觉计算模型基础上，提出了一

种应用于视频图像序列的注意焦点计算模型，利用对注意焦点的跟踪降低了计算量，更快地实现注意焦点的实时监测。但是 Itti 等人的视觉计算模型在特征的选择上还存在一定的局限性，选择更加有效的特征与描述形式，实现一个与视觉感知更接近的注意计算模型，是今后需要完善的地方。

#### 参考文献

- [1] 张 鹏, 王润生. 图像信息处理中的选择性注意机制研究[D]. 长沙: 国防科学技术大学, 2004.
- [2] Laurent I, Christof K, Ernst N. 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.
- [3] Laurent I, Christof K. A Comparison of Feature Combination Strategies for Saliency-based Visual Attention[C]//Proc. of SPIE. San Jose, CA, USA: [s. n.], 1999: 473-482.
- [4] Laurent I, Gold C, Koch C. Visual Attention and Target Detection in Cluttered Natural Scenes[J]. Optical Engineering, 2001, 40(9): 1784-1793.
- [5] Tsotsos J K, Culhane S M, Wai W Y K, et al. Modeling Visual via Selective Tuning[J]. Artificial Intelligence, 1995, 78(1/2): 507-545.
- [6] Comaniciu D, Ramesh V, Meer P. Kernel-based Object Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
- [7] Thracker N A, Ahere F J, Rocket P I. The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data[J]. Kybernetika, 1998, 34(4): 363-368.

(上接第 240 页)

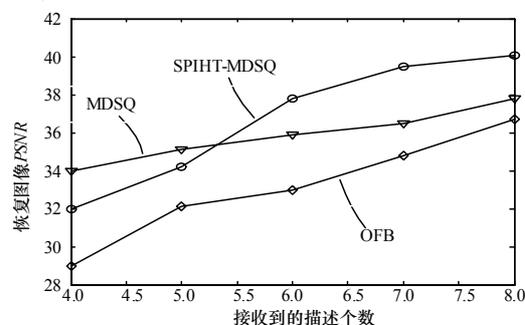


图5 接收到不同个数描述的恢复图像 PSNR



图6 接收到 8 个、6 个、4 个和 1 个描述下的接收端恢复图像

#### 5 结束语

多描述标量量化是一种简单高效的多描述编码形式，同时高效率的熵编码可以大大提高图像的编码效率。本文将多描述标量量化与 SPIHT 编码联合起来。首先将两描述的多描述标量量化进行多维度索引分配扩展，得到由粗量化索引形成的多个描述，然后采用 SPIHT 编码方法对每个多描述码流进行熵编码。实验结果表明，用高维多描述标量量化器产生多描述码流，再对每一个码流采用 SPIHT 进行熵编码，可以

获得很好的图像传输差错鲁棒性。本文的多描述编码方法尤其适合差错信道(特别是信道条件恶劣的无线网络环境)中的图像传输，能够在较大的信道差错率和严格的带宽限制下，获得较好的接收端恢复图像。

#### 参考文献

- [1] Vaishampayan V. Design of Multiple Description Scalar Quantizers[J]. IEEE Transactions on Information Theory, 1993, 39(3): 821.
- [2] Servetto S, Ramchandran K, Vaishampayan V. Multiple Description Wavelet Based Image Coding[J]. IEEE Transactions on Image Processing, 2000, 9(5): 813-826.
- [3] Wang Yao, Orchard M T, Reibman A R. Multiple Description Image Coding for Noisy Channels by Pairing Transform Coefficients[C]// Proc. of the 1st IEEE Workshop on Multimedia Signal. Princeton, New Jersey, USA: [s. n.], 1997: 419-424.
- [4] Shapiro M. Embedded Image Coding Using Zero-trees of Wavelet Coefficients[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3445-3462.
- [5] Said A, Pearlman W A. A New, Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 243-250.
- [6] Motwani R, Guillemot C. Tree-structured Over-sampled Filter-banks as Joint Source-channel Codes: Application to Image Transmission over Erasure Channels[J]. IEEE Transactions on Signal Processing, 2004, 52(9): 2584-2599.