doi: 10.3969/j.issn.1007-2861.2012.04.004

基于图像稀疏表示视觉显著度计算的自适应尺度调整方法

韩 彪, 杨卫英, 郑玉婷

(上海大学 影视艺术技术学院,上海 200072)

摘要: 图像适应是指将图像按照不同的纵横比进行拉伸和压缩,以适应不同显示比例的显示终端(如手机、掌上电脑(personal digital assistant,PDA)、投影仪)的一种技术.研究基于视觉显著度的图像适应,具体实现分为视觉显著度的计算和图像的加权拉伸与压缩两个步骤.采用加权的稀疏表达残差来表示视觉显著度.在图像的加权拉伸与压缩中,提出一种对横向与纵向使用不同加权变化的快速算法,在保证图像处理效果的前提下,达到快速图像适应处理的效果.仿真实验结果表明,该方法对于实验图像具有鲁棒性和有效性.

关键词: 图像适应;视觉显著度;加权拉伸

中图分类号: TP 391.4 **文献标志码**: A **文章编号**: 1007-2861(2012)04-0349-06

Image Retargeting Based on Visual Saliency Using Sparse Coding

HAN Biao, YANG Wei-ying, ZHENG Yu-ting

(School of Film and TV Arts & Technology, Shanghai University, Shanghai 200072, China)

Abstract: Image retargeting is a technique of adapting images to different aspect ratio for various display terminals such as cellphone, personal digital assistant (PDA) and projector. This paper proposes a method to deal with the problem using visual saliency. The method includes two steps: computing a saliency map, and modifying the image to suit different aspect ratio. A novel visual saliency model is used. The model indicates the weighted sparse coding residual as visual saliency. In the step of weighted resizing, a fast algorithm is proposed. It is a fast image adaptive process with no obvious distortion. The experimental results indicate the robust and effectiveness of the proposed method with different images.

Key words: image retargeting; visual saliency; weighted resize

随着数码相机等数字图像捕捉设备的普及,高分辨率的数字图像已经在人们的日常生活中随处可见.出于分享信息的需要,这些高分辨率的数字图像需要在不同的显示设备中进行显示.由于数字图像捕捉设备往往具有特定的捕捉长宽比,而手机、掌上电脑(personal digital assistant,PDA)、投影仪等显示设备的长宽比往往与其不尽相同,这就使得图像的捕捉和显示之间产生了一定的矛盾.为了解决这种矛盾,研究人员提出了使图像能自适应不同长宽比的方法,即图像适应方法.

图像适应的传统解决方法主要有两种:一种是按照显示比例对图像进行非等比拉伸;另一种则是

收稿日期:2011-05-11
基金项目:上海大学创新基金资助项目(SHUCX112306)
通信作者:杨卫英(1957~),女,副教授,研究方向为数字媒体技术. E-mail:yangweiying@staff.shu.edu.cn

按图像比例进行等比缩放,依据显示大小截取中心部分. 然而,这两种解决方法都存在一定的问题. 使用直接拉伸的方法,会使所获得的适应图像出现比例失真,特别是其中的重要物体(即所关注的物体)会产生严重的扭曲. 使用截取的方法,则会出现图像信息的丢失,特别是当重要物体处于画面边角的时候,该方法往往会丢失或改变图像所传递的信息.

现有的图像适应主要分为两个步骤:第一步是通过视觉关注度(visual attention)来定义图像中的关注程度或视觉显著度;第二步是根据视觉关注度的大小对图像进行变形.

对于视觉关注度的计算,现有的主要方法有 Itti 算法[1]、频谱残差(spectrum residual,SR)算法[2]和 Judd 算法[3]. 一般来说,视觉关注分为两个部分,即从底向上(bottom-up)的关注和从顶往下(top-down)的关注[4]. 从底向上的关注是指我们所获取的信息都是来源于图像或者视觉刺激本身,这种关注过程将这样的信息直接转化为对视觉的关注. 从顶往下的关注是指有意识参与的关注,即通过意识的显性控制或者由先验知识参与控制的方式,将关注转移到特定的地点,通过和眼动仪数据进行比较,就可以大致得到该算法与人眼相关机能的相似性. Itti 算法是视觉关注领域最经典的算法之一,它是由 Itti 等[1]在 1998 年提出的一个视觉关注模型. Itti 算法忠实地描述了特征综合理论,并使用颜色、方向和亮度作为特征来衡量显著性. SR 算法是由上海交通大学的 Hou 等[2]在 2007 年提出的. SR 算法认为,显著性区域可以通过输入的视觉刺激与频域 Log 谱的先验知识的残差来表示. Judd 算法由美国麻省理工学院的 Judd 等[3]在 2009 年提出,是一种通过机器学习的方法来预测视觉关注的算法. Judd 算法使用支持向量机的方法对大量眼动仪数据进行机器学习,从而得到最终结果. 通过这些算法得到的结果与人眼眼动的实验数据进行比较,误差都相对较大,对于图像适应的应用前景有限.

与传统图像适应方法类似,解决图像变形的方法主要有图像裁剪和变形. Santella 等[5]提出的根据图像视觉显著性的裁剪方法,仅仅保留了图像中用户关注的区域,而裁剪了周围的图像内容. 这样的方法会完全丢弃图像中大量的背景信息,不能保证图像信息的完整性. 基于变形的方法主要分为两类:一类是基于拉伸压缩,如利用网格[6]、前景分割[7]的方法;另一类是基于抽丝[8]的方法. 基于网格的拉伸算法首先建立图像的参数化网格,再通过这样的网格变化对图像进行拉伸. 前景分割是先分割出前景,再将前景融合到拉伸后的背景中去. 基于抽丝是先计算出图像中能量最小的裂缝,然后抽去. 现有的这些方法都存在数据运算量大、算法鲁棒性不足的问题.

针对以上方法存在的不足,本研究提出了一种基于视觉显著度的图像适应方法,从两个方面对图像适应方法进行了改进. 首先,使用一种新的视觉显著度的度量方式,即使用加权的稀疏表达残差作为显著度的度量方式,该方式与人眼眼动相比,准确率更高,预测效果更好;其次,使用一种快速的加权拉伸压缩算法,直接利用视觉显著度图中的信息进行应用,该方法的数据运算量较小,并且具有较强的鲁棒性.

1 视觉显著度计算

为了确认人眼所关注的图像区域,需要计算图像的视觉显著度,并通过与眼动数据库的比较,来判断得到的视觉显著度的好坏程度. 所谓眼动数据库,就是使用眼动仪采集得到的人眼的眼动数据,也就是真实人眼观察图像所关注的地点的统计. 通过绘制接受者操作特征(receiver operating characteristic,ROC)曲线,并计算曲线下的面积大小就可以作为视觉显著度好坏的评判标准.

视觉显著度的计算是计算机视觉领域和计算神经科学的一个重要研究方向. 本研究使用了一种加权的稀疏表达残差算法来表示视觉显著度,并通过以下两个步骤得到视觉显著性图:① 图像的稀疏表达;② 加权的稀疏表达残差的计算.

1.1 图像的稀疏表达

图像的稀疏表达最初由 Olshausen[9]提出,是一种人类早期视觉系统初级视觉皮层中简单细胞的模型. 该模型可以在一定程度上模拟人脑中处理视觉信息的过程. 这个观念在推广到计算机视觉领域后,在数学家和统计学研究者的共同关注下,稀疏表达的概念得到了很好的推广,且作为一个数学问题的稀疏表达,也有了相应的优化解决方法. 稀疏表达是一种最小熵编码[9],得到的熵是整个图像中最小的部分. 当对稀疏表达进行还原时,其还原出的图像也是图像中熵最小的部分,因此,通过计算原图像和还原图像的差值,就可得到一个最大熵的部分. 研究表明,人们视觉关注的区域应该是一个图像中熵最大的区域,因此,原图像和还原图像的差值区域就是视

觉关注的区域[10].

稀疏表达的第一步就是将图像分为大小相同的块,即图像 $X = \{x_1, x_2, \cdots, x_n\}$,其中 n 为图像块的数量. 每个图像块可以由一组稀疏基(稀疏编码)和字典的线性乘积得到,即

$$x_i = Da_i + r_i, \tag{1}$$

式中,D 为稀疏表达中的字典,a_i 为稀疏表达编码,r_i 为稀疏表达结果与原图像的差值,也就是稀疏表达残差. 所谓稀疏表达就是在求最稀疏的情况下,稀疏表达残差最小. 一个矩阵的稀疏性可以由其 0 阶范数来表示,因此,式(1)中的约束条件为

$$\min \lambda \|a_i\|_0 + \|r_i\|_2, \tag{2}$$

式中,λ 为平衡稀疏性和数据完整性(残差最小)的参数,但这个优化问题也是很难解决的.

2006 年,Donoho[11] 证明了对于大部分系统而言,求最稀疏的优化问题可以由其 1 阶范数的解来近似表示,即

$$\min \lambda \|a_i\|_1 + \|r_i\|_2. \tag{3}$$

该 1 阶范数优化问题也就是经典的 Lasso 线性回归问题,可以使用最小角度回归(least angle regression, LARS)算法[12] 解决. LARS 算法是一种解决稀疏表达问题的放松算法,也是解决这类问题的常用算法. 通过解决该优化问题,就可以得到式(1)中各个参数的值,也就可以得到稀疏表达残差.

1.2　加权的稀疏表达残差的计算

对于使用稀疏表达进行视觉显著度计算这一特定的应用,本研究提出了利用加权的稀疏表达残差的方法来表示图像的视觉显著度. 通过计算每一个图像块的稀疏表达残差和稀疏表达编码的乘积,可以得到每一个图像块的视觉显著性图,通过将这些视觉显著性图进行组合,就得到了整个图像的视觉显著性图.

以上整个稀疏表达残差的计算过程中,尚没有解决字典的训练. 本研究使用上海交通大学 Li 等[13] 提出的局部字典的方法(该方法认为任意一个图像块的字典是其周边重叠的图像块),可以快速地训练出字典,得到的稀疏表达编码可以作为图像块的奇异程度的度量. 该编码在本研究中作为稀疏表达残差的加权.

图 1 为本研究部分视觉显著性图的实验结果,并且与 Itti 算法、SR 算法和 Judd 算法所得到的显著性图进行了比较.

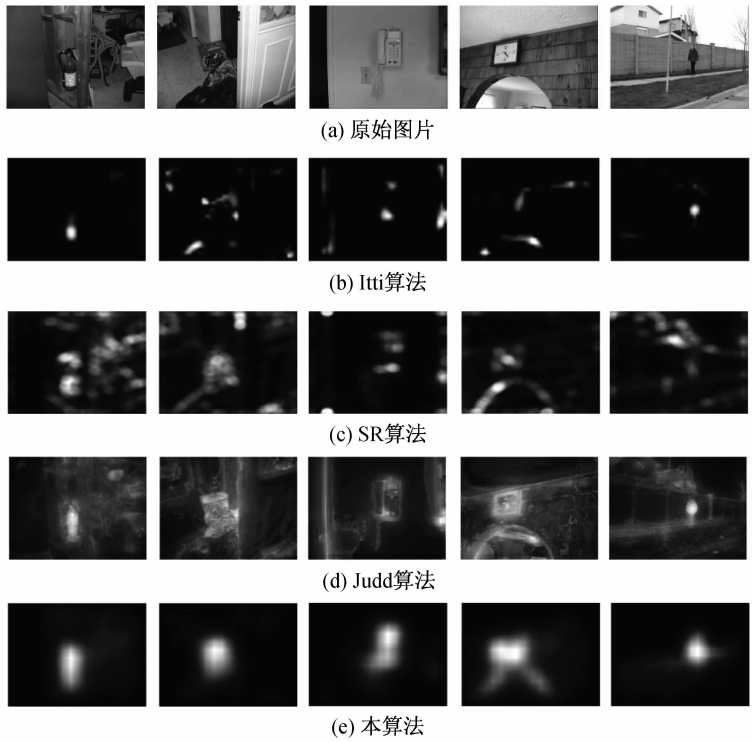

图 1　部分视觉显著性图的实验结果

Fig. 1　Some experimental results of different algorithms

通过与 Bruce 等[10]提供的眼动数据库进行对比,本研究计算了不同算法与人眼真实眼动的相似程度. 表 1 为 120 幅图片的眼动数据库 ROC 下面积与其他 3 种算法的对比结果,其值越大越好.

表 1 ROC 曲线下区域面积对比

Table 1 Area under the curve (AUC) of ROC curve %

	Itti 算法	SR 算法	Judd 算法	本算法
AUC 百分比	58.69	71.72	77.95	82.64

根据表 1 的结果,本研究对这 4 种算法作如下分析. 首先,对于 Itti 算法,其本质是对特征综合理论的一种实现,通过对各种人为提取的"特征"进行加权综合,得到最终的结果. 这种算法成立的前提是人脑内部通过这样的"特征综合理论"实现关注,但这种理论的正确性值得探讨. 其次,SR 算法和 Judd 算法分别使用了频谱和机器学习的方式进行视觉显著度的计算. 这种计算方式虽然可以得到相对较好的结果,但其本身并没有对视觉关注的本质进行探讨,因此,并不能完全反映图像中的显著性. 最后,对于本视觉关注算法,其基础是稀疏表达理论. 该理论是对人脑较为深层次的模拟,因此,本算法可能更接近人脑中发生视觉关注的本质. 实验结果也证明,加权的稀疏表达残差的模型在表现上比其他模型更能表达图像中的显著性.

2 图像适应变形方法

在计算出图像的视觉显著性图之后,对图像进行图像适应变形以实现图像适应. 本研究使用了一种快速图像适应变形的方法,其流程如图 2 所示.

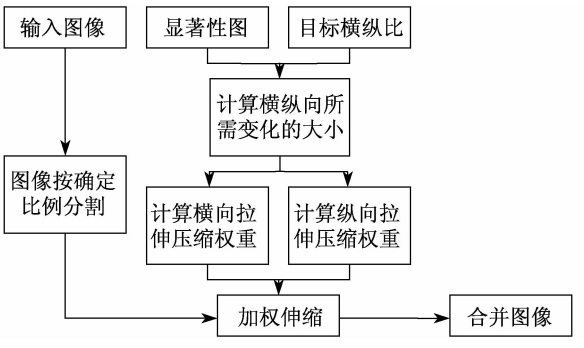

图 2 基于视觉显著性图的图像适应变形方法

Fig. 2 Image retargeting method based on visual saliency

2.1 基于图像视觉显著性图的适应变形

图 2 所示的图像适应变形方法的步骤如下:① 将输入图像按照确定比例分割为横向和纵向的图像块;② 通过输入显著性图和目标横纵比,计算出横纵向所需的拉伸;③ 分别计算出横向和纵向拉伸压缩加权的权重,对图像块进行加权伸缩;④ 通过合并这些伸缩结果,得到最终的伸缩图像.

首先,根据已确定的 e 个像素为一个单位,将输入图像按横向和纵向分为$\frac{h}{e}$和$\frac{w}{e}$个长条形和纵条形图像块,其中 h 和 w 分别为图像的高度和宽度,这些图像块就是进行适应变形的基本单位. 在本研究中,e 取 10 像素.

接着,对显著性图横向和纵向求和,得到两个显著性加权矩阵,通过统计这两个矩阵小于某一阈值的个数,确定横向和纵向拉伸压缩所要附加的权重. 此过程可表示为

$$\frac{h+\Delta h}{w+\Delta w}=\frac{h+k\cdot i}{w+k\cdot j}=\rho, \tag{4}$$

式中,Δh 为高度变化的量,Δw 为宽度变化的量,i 和 j 分别为纵向和横向显著性加权矩阵中小于某一阈值的数量(在本研究中这一阈值的大小为 0.05),ρ 为目标的横纵比,k 为未知参数. 通过求解式(4)的一元一次方程,得到高度变化量和宽度变化量.

为计算横纵向变化,需分配到每一个图像变形的基本单位,即那些以 e 像素为单位的长条形和纵条形图像块,因此,要对显著性矩阵进行压缩. 在本研究中,首先将显著性矩阵进行归一化,然后使用最邻近法将显著性矩阵压缩到原大小的$\frac{1}{e}$,并通过该数值对每一个图像块进行加权的变形处理. 此过程可表示为

$$\begin{cases}\Delta y_m=\Delta h\cdot\dfrac{1-I_m}{\sum\limits_c(1-I_c)},\\ \Delta x_m=\Delta w\cdot\dfrac{1-J_m}{\sum\limits_c(1-J_c)},\end{cases} \tag{5}$$

式中,Δy_m 为第 m 个纵向像素块的变化量,Δx_m 为第 m 个横向像素块的变化量,I_m 为矩阵 $\boldsymbol{I}$ 中第 m 个元素的值,J_m 为矩阵 $\boldsymbol{J}$ 中第 m 个元素的值. 在得到横向和纵向的变化量后,对图像块进行拉伸和压缩,再将经过拉伸和压缩后的图像块合并在一起,得到最终的适应变形图像. 由于本方法只使用了拉伸和压缩操作,所以数据运算量小,鲁棒性强.

2.2 实验结果与分析

本研究在 Intel Core2 2.4 GHz CPU 的苹果电脑(MAC)上,使用 Matlab 对多种类型的图像进行了仿

真实验,部分结果如图 3 ~ 图 5 所示.

如图 3 所示,本研究分别进行了拉伸和压缩实验,将原比例为 4∶3的图像分别适应到了 2∶1和 3∶4 的显示比例. 可以明显地看出,使用本算法所得到的图像对于用户关注位置的失真更小,用户视觉体验更好,基本可以做到在没有明显拉伸和压缩的情况下,进行图像适应.

如图 4 所示,本研究分别使用了不同视觉显著性算法进行了图像适应的实验,将原比例为4∶3的图像适应到了 2∶1的显示比例. 结果发现,本算法具有较为明显的优势. 为了得到以上几种显著性算法的时间复杂度对比,本研究使用这些算法在 Bruce 等[10]提供的 120 幅图像上进行了测试,并统计了各算法的平均消耗时间,结果如表 2 所示.

图 3　与直接拉伸比较

Fig. 3　Comparison with resizing directly

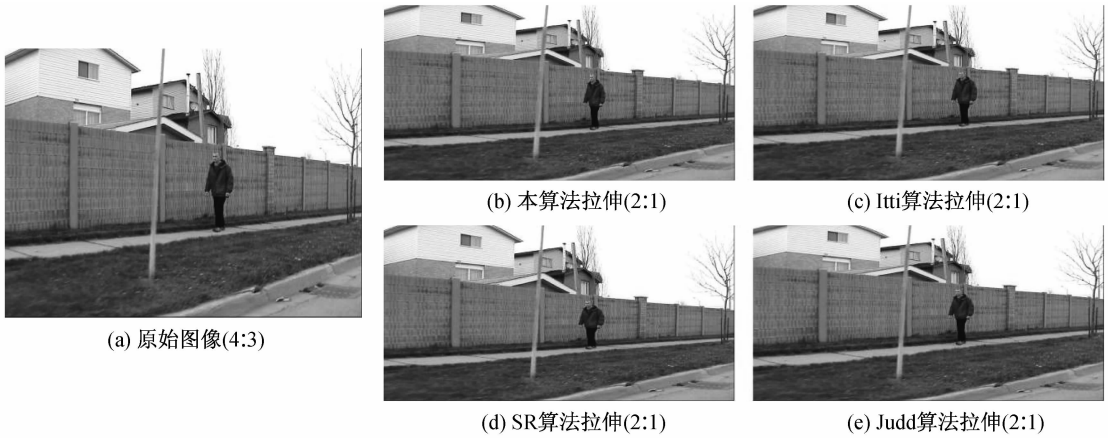

图 4　与其他视觉显著性算法进行图像适应比较

Fig. 4　Comparison with other saliency methods

由表 2 可见,由于 Itti 算法和 SR 算法没有使用如 Judd 算法中的机器学习方法,因此,其实现速度都相对较快. 而本算法由于使用了 Li 等[13]提出的局部字典方法和较为经典的Lasso问题解法,因此,实

(a) 原始图像(4:3) (b) 本算法压缩(3:4) (c) 直接拉伸(3:4) (d) 剪裁(3:4) (2) 抽丝(3:4)

图5 与其他图像适应方法比较

Fig. 5 Comparison with other image retargeting methods

现速度也相对较快.

表2 平均计算时间

Table 2 Average time used for saliency computing s

	Itti 算法	SR 算法	Judd 算法	本算法
时间	0.51	0.12	24.67	1.24

如图5所示,本研究将几类图像适应方法进行了对比. 可以看到,使用直接裁剪的方法会丢失很多背景信息,而使用抽丝的方法则会造成部分图像的扭曲变形,其鲁棒性表现不足. 本方法在尽可能保留图像信息和保持图像不失真的前提下,可以较好地实现图像适应.

3 结束语

本研究提出了一种基于视觉显著度的图像适应方法. 由于使用了较为准确的视觉显著度模型,因此,本方法对于用户视觉感知的表现良好. 在图像适应变形方面,本研究使用了一种快速鲁棒的方法,可以在平衡图像信息保存和图像失真的前提下,进行更好的图像适应.

在未来的研究中,将进一步尝试将这样的方法转移到时域中去,通过对视觉显著度的进一步研究来实现对视频信息的适应化处理.

参考文献:

[1] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.

[2] HOU X, ZHANG L. Saliency detection: a spectral residual approach [C]// The 20th IEEE Conference on Computer Vision and Pattern Recognition. 2007:1-8.

[3] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look [C]// 2009 International Conference on Computer Vision. 2009:2106-2113.

[4] YANTIS S. Control of visual attention [M]. London: Psychology Press, 1998:223-256.

[5] SANTELLA A, AGRAWALA M, DECARLO D, et al. Gaze-based interaction for semi-automatic photo cropping [C]// Proceedings of the 2006 Conference on Human Factors in Computing Systems. 2006:771-780.

[6] 时健,郭延文,杜振龙,等. 一种基于网格参数化的图像适应方法[J]. 软件学报,2008,19(Z1):19-30.

[7] SETLUR V, LECHNER T, NIENHAUS M, et al. Retargeting images and video for preserving information saliency [J]. IEEE Computer Graphics and Applications, 2007, 27(5):80-88.

[8] AVIAN S, SHAMIR A. Seam carving for content-aware image resizing [J]. ACM Transactions on Graphics, 2007, 26(3):267-276.

[9] OLSHAUSEN B. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381(6583):607-609.

[10] BRUCE N, TSOTSOS J. Saliency based on information maximization [C]// Advances in Neural Information Processing Systems. 2006:155-162.

[11] DONOHO D. For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution [J]. Communications on Pure and Applied Mathematics, 2006, 59(7):907-934.

[12] EFRON B, HASTIE T, JOHNSTONE I, et al. Least angle regression [J]. The Annals of Statistics, 2004, 32:407-499.

[13] LI Y, ZHOU Y, XU L, et al. Incremental sparse saliency detection [C]// The International Conference on Image Processing. 2009:3093-3096.