

基于区域关系图的图像显示适配技术*

任桐炜¹, 刘焱², 武港山¹⁺

¹(南京大学 计算机科学与技术系, 江苏 南京 210093)

²(香港理工大学 计算学系, 香港)

Image Retargeting Based on Region Relation Graph

REN Tong-Wei¹, LIU Yan², WU Gang-Shan¹⁺

¹(Department of Computer Science and Technology, Nanjing University, Nanjing 210093, China)

²(Department of Computing, Hong Kong Polytechnic University, Hong Kong, China)

+ Corresponding author: E-mail: gswu@nju.edu.cn

Ren TW, Liu Y, Wu GS. Image retargeting based on region relation graph. *Journal of Software*, 2010,21(9): 2237-2249. <http://www.jos.org.cn/1000-9825/3850.htm>

Abstract: Image retargeting techniques aim to adapt images to the target screens with different sizes and aspect ratios. Unfortunately, image retargeting for mobile device screens suffers from the problems of limited display area and various aspect ratios. This has been one of the hot topics in multimedia research. This paper proposes a novel image retargeting approach based on region relation graph. First, energy map and sensitivity energy maps are calculated by visual attention and weighted gradient, respectively. Then, the original image is decomposed into curve-edge trapezoid meshes, further represented by region relation graph. Based on the region relation graph, image retargeting is formulated as a quadratic programming problem constrained by energy maps, which emphasizes the important image parts and reduces the visual distortion by optimally relocating the key mesh vertexes. Finally, the target image is generated based on the optimal solution. Experimental results show the effective and efficiency of the proposed approach.

Key words: image retargeting; region relation graph; multiple energy map constraint; curve-edge trapezoid mesh representation; preliminary scaling

摘要: 图像显示适配技术用于改编图像,以适应在不同尺寸和长宽比的目标屏幕上展示。其中,面向移动设备的图像显示适配面临展示面积受限、长宽比不固定等问题,已经成为多媒体领域的研究热点之一。针对现有方法的缺陷,提出了一种基于区域关系图的图像显示适配方法。首先,根据视觉关注程度和加权梯度分别计算出图像的重要度能量模板和敏感度能量模板。然后,将原始图像分解为曲边梯形网格,并进一步表示为区域关系图。在区域关系图基础上,图像显示适配被规约为用能量模板约束的二次规划问题,通过优化关键网格顶点重定位来生成重要物体突出且视觉扭曲不明显的目标图像。实验结果表明,该方法在图像显示适配中可以取得较好的效果。

关键词: 图像显示适配;区域关系图;多能量模板约束;曲边梯形网格表示;预缩放

* Supported by the National Natural Science Foundation of China under Grant Nos.60533080, 60721002, 60975043 (国家自然科学基金); the Hong Kong General Research Fund of China under Grant No.PolyU 5204/09E (香港研资局优配研究金)

Received 2009-09-13; Accepted 2010-03-11

中图法分类号: TP301

文献标识码: A

多媒体技术的发展使得数字图像成为在手机、PDA 等移动设备间传递信息的重要媒介。然而,相对于尺寸较小且形状各异的移动设备屏幕,现有的图像信息大多具有过高的分辨率和限定的长宽比,需要经过改编才能更好地展示。自动改编图像以适应不同尺寸和长宽比目标屏幕的技术通常被称为“图像显示适配技术(image retargeting)”。面向传统展示设备(如电视或计算机屏幕)的图像显示适配通常采用均匀缩放技术。但这类技术用于面向移动设备的图像显示适配时,会造成重要物体过小而无法识别以及存在严重视觉扭曲等问题。因此,如何保证改编后的图像在面积受限、长宽比不固定的移动设备屏幕上展示时,既能突出重要物体和反映原图基本信息又满足视觉扭曲不明显的要求,已经成为多媒体领域广受关注的问题。

1 相关工作

为了克服均匀缩放技术的缺陷,现有的图像显示适配技术采用了“内容相关(content-aware)”的思想,即根据图像中各部分内容重要程度的差异来确定其取舍或缩放程度,从而达到减少信息损失的目的。根据处理层次的不同,现有方法大致可以分为3类:对象层次的显示适配技术(object level retargeting)、像素层次的显示适配技术(pixel level retargeting)和区域层次的显示适配技术(region level retargeting)。

对象层次的显示适配技术将原始图像准确区分为重要物体与背景两部分,通过对两部分的不同处理来达到突出重要物体的目的。其中,裁剪(cropping)方法^[1-3]检测出包含尽可能多重要物体且与目标屏幕长宽比相同的最小区域,并通过均匀缩放该区域来生成目标图像。这类方法会造成背景信息的严重丢失,不利于对图像内容的理解,且无法处理物体之间的背景部分,对于多个分散物体的情形不适用。部分研究者通过裁剪生成原始图像中多个重要区域并按序展示^[4,5],该策略可以有效地解决上述问题,但不适用于要求结果为单幅图像的应用。针对裁剪方法的缺陷,基于分割的重构(segmentation based recomposition)方法^[6]将原始图像分割为重要物体和背景,并将填充好且缩放至目标尺寸的背景与分割出的物体重新组合。然而,基于分割的重构方法依赖于物体分割、背景填充等技术的效果,还可能造成物体空间关系和尺寸比例的失真。除上述缺点外,对象层次的显示适配技术建立在能够将物体与背景准确区分的基础上,这通常很难自动实现。

像素层次的显示适配技术将图像中的每个像素看作一个处理单元,并对每个像素赋给一定的能量值,通过求解带约束的优化方程来在目标图像中突出高能量的像素。其中,抽丝(seam carving)方法^[7,8]定义了一种称为“丝(seam)”的8-连通像素序列,并通过动态规划算法迭代地去掉能量最低的“丝”。非均匀重定位(non-homogeneous relocation)方法^[9,10]将显示适配规约为线性规划问题,通过求解每个像素在目标图像中的位置来实现图像显示适配。像素层次的显示适配技术克服了对象层次方法对预处理要求过高的缺陷;但它们通常需要较大的计算开销,并且很难通过约束条件来有效地保持图像的结构信息。

区域层次的显示适配技术将图像划分为一系列的区域,并将每个区域作为一个处理单元,通过优化各个区域的变形来获取目标图像。与对象层次的方法相比,区域层次的显示适配方法用区域划分代替物体分割,降低了对预处理的要求;与像素层次的方法相比,区域层次的显示适配方法用区域代替单个像素作为处理单元,可以减少计算开销和更好地保持图像结构。因此,在区域层次进行图像显示适配已成为当前研究的热点,本文方法也采用了这一思路。在现有的区域层次方法中,基于鱼眼变形(fisheye-view warping)的方法^[11]根据预先选定的模板进行区域划分和区域变形。该方法需要人工选择模板,但在图像结构复杂时容易造成视觉扭曲。基于网格参数化(mesh parametrization)的方法^[12]将图像用内容相关的三角网格表示,通过均匀缩放可能包含重要物体的网格和优化变形只包含背景的网格来生成目标图像。该方法需要预先判定对网格是否包含重要物体,并且会在紧靠重要物体的区域造成严重扭曲。基于缩放-拖拽(scale-and-stretch)的方法^[13]将图像表示为均匀的方形网格,通过减少高能量区域的变形来消除视觉扭曲。该方法着眼于消除目标图像中的视觉扭曲,没有考虑突出重要物体。

针对现有区域层次方法中存在的不足,本文在前期工作^[14,15]的基础上,提出了一种新颖的基于区域关系图的图像显示适配方法。与现有方法相比,本文的贡献主要包括:(1) 提出了多能量模板约束策略,有效地克服了单

个能量模板无法同时约束适配效果不同方面的缺陷,并讨论了多能量模板约束策略对现有其他方法的适用性;(2)提出了一种曲边梯形网格表示法及根据图像的网格表示构建区域关系图的方法,该方法能够更好地描述图像结构信息;(3)提出了一种基于区域关系图的图像显示适配方法,并将突出重要物体和消除视觉扭曲两个适配目标统一在优化方程中,取得了更好的显示适配效果。

图1展示了本文方法的基本思路:首先,计算出原始图像(如图1(a)所示)的能量模板(如图1(b)所示),用于表示图像内容在重要度、敏感度等方面的性质;接着,根据区域划分结果将原始图像表示为曲边梯形网格(如图1(c)所示),并进一步构建区域关系图(如图1(d)所示);然后,在能量模板的约束下,根据区域关系图对网格变形进行优化(如图1(e)所示),重定位关键网格顶点;最后,根据关键网格顶点重定位的结果,生成目标图像(如图1(f)所示)。

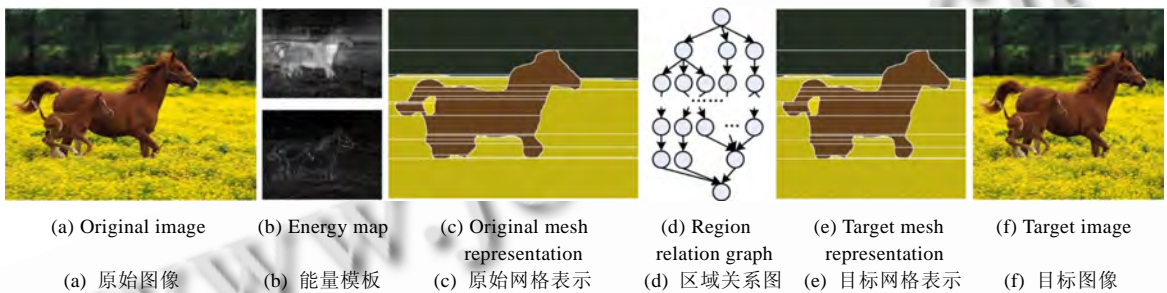


Fig.1 An overview of the proposed approach

图1 本文方法的基本思路

2 多能量模板约束策略

图像内容的各种性质分别在不同方面影响着图像显示适配的效果.例如,内容的重要程度影响着图像内容的取舍,对扭曲的敏感程度决定了图像内容允许变形的程度.在图像显示适配技术中,重要度、敏感度等性质通常被统称为“能量(energy)”,用 $[0,1]$ 中的值来表示,能量值越大,表明相应的性质越显著.根据每个像素对应的能量值,图像的每种性质可以表示为一个大小为 $W \times H$ (W 和 H 分别为图像的宽和高)的矩阵,该矩阵称为“能量模板(energy map)”.同一幅图像通常对应多个能量模板,用于反映图像内容的不同性质.

2.1 多能量模板约束策略的优势

现有的显示适配方法都只采用单一的能量模板来约束图像显示适配.单能量模板约束策略没有全面地考虑图像内容的各种性质,可能会给未约束的方面带来问题.以抽丝方法为例,它选择以梯度值作为能量来保持物体边缘和减少视觉扭曲,没有对重要内容保持方面进行约束.这使得一旦重要物体的边缘被破坏,物体的结构往往由于内部梯度值不高而被严重损坏^[9].如果简单通过加权和、乘积等方式将约束不同方面的能量融合到同一个能量模板中,则会混淆各种能量的作用,无法很好地改进适配效果.例如,当用内容重要度和梯度的加权和作为能量时,抽丝方法虽然可以较好地突出重要物体,但可能由于梯度在能量值计算中的权重被削弱而造成背景中强边缘的严重扭曲^[12].

针对单一能量模板约束的缺陷,我们提出了一种采用多个能量模板来分别约束图像显示适配中不同方面的策略.在多能量模板约束策略中,对于优化方程中的同一个变量,会根据与目标函数或约束条件相关联的图像内容性质而选取相应的能量来进行约束.这一策略在不混淆不同能量作用的情况下,全面考虑了图像内容的各种性质,因而有助于取得更好的显示适配效果.在本文方法中,我们使用了两种能量模板:重要度模板和敏感度模板.更多种类的能量模板可以根据需要进行添加,以约束图像显示适配中的其他方面.图2展示了使用不同能量模板生成的适配效果.可以发现:当仅使用重要度模板时,目标图像中存在明显的视觉扭曲(如图2(c)所示);当仅使用敏感度模板时,目标图像中的重要物体没有被突出(如图2(d)所示);而采用多能量模板约束策略时,目标

图像在突出重要物体和消除视觉扭曲方面均取得了较好的效果(如图 2(e)所示).



Fig.2 Energy maps and the results using different energy map(s)

图 2 能量模板和使用不同能量模板生成的适配效果

2.2 重要度模板和敏感度模板的计算

重要度模板用于刻画图像内容的重要程度,即该内容包含图像信息的多少.由于目前还不能完全从语义层面来描述图像内容的重要度,仅现有的图像显示适配方法通常采用基于低层视觉特征的显著度模型^[1]和人脸等重要物体^[11]来共同表示图像内容是否被关注.在本文方法中,我们首先采用 Itti 等人提出的显著度模型^[16]来检测视觉上容易引起关注的图像区域.在此基础上,再采用 Viola 等人提出的人脸检测算法^[17]检测出人脸区域,并将其能量值设为最大值 1.图 2(a)展示了重要度模板的示例,从白色到黑色表示能量由最高到最低.

敏感度模板用于刻画图像内容对变形的敏感程度,即该部分内容变形后是否容易被肉眼察觉.通常认为,高梯度区域发生变形更加容易引起人眼的察觉^[7];此外,敏感度还与该区域的内容重要度^[13]、位置^[1]等因素有关.因此,在本文中,我们采用加权梯度来描述图像内容的敏感度.我们以内容重要度能量最大的区域为中心(区域划分将在第 3.1 节讨论),以各中心的平均重要度能量为峰值建立混合高斯模型,得到各点的梯度权重,并结合梯度值计算出各点的敏感度能量.值得注意的是,由于图像在水平方向和垂直方向的梯度值不同,我们需要计算不同的敏感度模板来分别约束水平方向和垂直方向的图像显示适配.图 2(b)展示了用于约束水平方向适配的敏感度模板示例,从白色到黑色表示能量由最高到最低.

2.3 多能量模板约束策略的推广应用

多能量模板约束策略除有助于本文方法取得更好的适配效果外,还可以用于改进现有的其他图像显示适配方法.在抽丝方法中,可以在使用梯度作为能量来消除视觉扭曲的同时,根据每个区域的重要度确定允许去除像素数量的上限,以避免重要物体的丢失.在非均匀重定位方法中,可以在目标函数中使用重要度能量来突出重要物体,而在约束条件中使用敏感度能量来减少视觉扭曲的产生.在基于网格参数化的方法中,可以利用重要度模板来确定可能包含重要物体的网格,而在参数化过程中利用敏感度模板来约束每个网格的变形程度.在基于缩放-拖拽的方法中,可以在使用敏感度模板约束网格变形程度的同时,利用重要度模板来约束网格的大小.

由此可见,现有基于优化求解的图像显示适配方法均可采用多能量模板约束策略来改进适配效果,达到在突出重要物体的同时又能够消除明显视觉扭曲的目的.多能量模板约束策略具有普遍的适用性.

3 基于曲边梯形网格的区域关系图表示

区域层次的显示适配方法通常使用网格来表示图像的区域划分,并通过优化根据网格间关系及能量模板进行约束的网格变形来实现图像的显示适配.因此,网格表示及网格间关系描述方法的优劣直接影响到显示适配的效果与性能.

3.1 图像的曲边梯形网格表示

现有显示适配方法中使用的是基于内容相关的三角网格表示^[12]或均匀方形网格表示^[13].这两种网格表示方法都将区域边缘分解成小段来近似表示,不利于从整体上保持区域边缘的形态;此外,这两种表示方法会在均质区域内形成大量网格顶点,从而造成优化求解中计算开销的增加.

针对这些问题,本文提出了一种根据均质区域边缘构建网格的曲边梯形网格表示方法(curve-edge trapezoid mesh representation).为了简化讨论,下文中我们仅仅讨论面向水平方向适配的网格表示,面向垂直方向适配的网格表示可以通过类似的方法生成.

首先采用 Mean-Shift 算法^[18]将原始图像(如图 3(a)所示)划分成均质的区域(如图 3(b)所示).接着,我们检测出每个区域的边缘和区域内长度超过给定阈值的近似直线边缘,并将区域边缘分解为其上各点 y 坐标保持单调变化的曲线段.对于每条曲线段(包括区域内的近似直线边缘),如图 3(c)所示,我们选取其顶点和 x 坐标局部极值点(局部最左或最右的点)作为关键点(图 3(c)中的圆形标记和星形标记).这样,位于同一曲线段上的关键点序列便构成了对该曲线段形态的近似描述.对于位于图像内部的关键点(圆形标记),我们在其左右两侧同一行的像素中各选取一个辅助点,辅助点位于离该关键点最近的曲线段(三角标记)或者图像边缘上(方形标记).对于位于图像边缘上的关键点(星形标记),可以类似地选取 0 或一个辅助点.根据关键点与辅助点之间的连线及曲线段,可以构造出图像的曲边梯形网格表示(如图 3(d)所示).

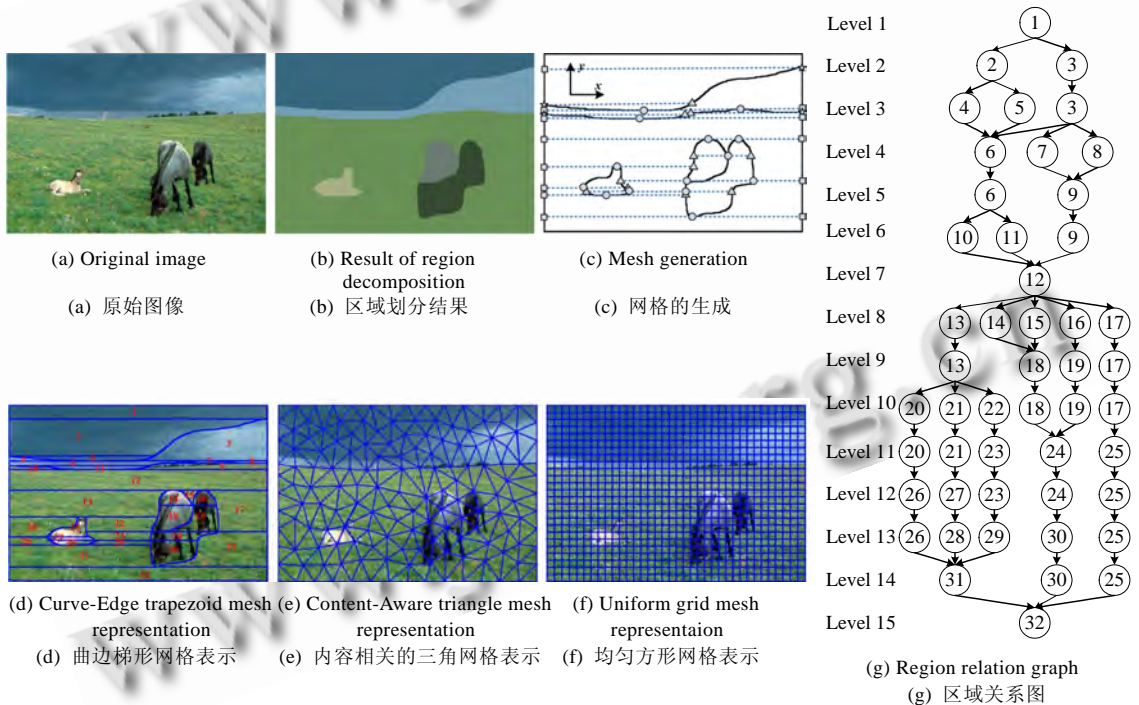


Fig.3 Curve-Edge trapezoid mesh representation and region relation graph

图 3 曲边梯形网格表示和区域关系图

直接运用上述方法会因区域边缘不平滑或图像构成复杂而生成较多的网格,不利于进一步的计算求解.我们预先采用轮廓演化的方法对区域边缘进行平滑^[19],并将像素数量小于指定阈值(本文实验中为像素总量的 1/50)的网格和与其共边且两者平均重要度能量和平均敏感度能量的差别均小于指定阈值(本文实验中为 0.1)的网格合并.实验表明,该策略可以有效地解决网格数量过多的问题.

图 3(d)~图 3(f)展示了图 3(a)的不同网格表示(图 3(e)源自文献[12]).与其他网格表示方法相比,曲边梯形网

格表示法直接利用均质区域来生成网格,并使用关键点来描述网格边缘,有助于在网格变形中更好地保持边缘形态.此外,曲边梯形网格表示法避免了在均质区域内产生大量的网格结点,且生成的网格数量仅与图像结构的复杂程度有关,而与图像尺寸大小无关.因此,曲边梯形网格表示法生成的网格数量通常会远远小于现有的方法,尤其是在图像尺寸较大且结构较简单的时候.这会大大减少进一步求解所需的计算量.以图3为例,使用内容相关的三角网格表示和均匀方形网格表示时所产生的网格数量分别为261个和900个,而使用曲边梯形网格表示时产生的网格数量仅为32个.

3.2 区域关系图的构建

在显示适配过程中,各网格的变形之间存在着相互影响.例如,各网格在变形后处于同一行的宽度之和应等于目标图像宽度.此外,本文方法中部分网格的顶点位置需要根据相邻网格的变形结果来确定(具体请参见第4.2节).因此,我们需要对网格之间的空间位置关系进行描述.在本文方法中,我们采用了区域关系图(region relation graph)的表示方法将每个网格作为一个结点,通过构建有向图来刻画网格间的空间关系.

与曲边梯形网格表示一样,区域关系图的构建也与显示适配的方向有关.以构建用于水平方向适配的区域关系图为例,我们按上底边出现的次序对网格进行编号(图3(d)),并依次扫描图像各行中从左到右网格出现的序列.如果当前行网格出现的序列与上一行相同,则扫描下一行;否则,根据当前行网格出现的序列,在区域关系图中新增一层结点,每个结点用网格的编号标识.在新增结点时,如果相同编号的结点在上一层中也存在,则添加一条由上一层同号结点指向该结点的边;否则,在区域关系图的上一层结点中找到那些对应网格的下底边与该结点对应网格上底边具有重合部分的结点,并添加由这些结点指向该新增结点的边.这样,图像中网格间的关系可以表示为如图3(g)所示的区域关系图.区域关系图能够很好地描述网格之间的空间关系.例如,区域关系图中位于同一层的结点表示在图像中某一行同时出现的网格序列;同层相邻结点对应的网格在水平方向相邻;父结点和子结点对应的网格在竖直方向相邻,且父结点的下底边(之和)等于子结点的上底边(之和).

4 基于区域关系图的图像显示适配

根据图像的区域关系图表示,图像显示适配转化为求解满足约束条件的最优网格变形方法.在现有的图像显示适配方法中,网格变形通常采用重定位网格顶点方法来实现.因此,图像显示适配可以看作是网格顶点重定位的带约束优化问题.

4.1 单个网格的变形约束

图4展示了一个水平方向网格变形的示例,竖直方向的网格变形可以采用类似的方法.

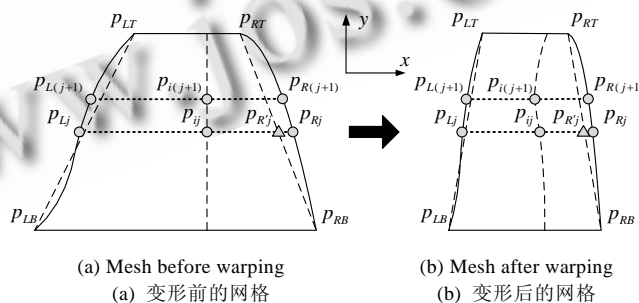


Fig.4 Horizontal warping of curve-edge trapezoid mesh

图4 曲边梯形网格的水平方向变形

在下面的讨论中,我们用 (x_{ij}, y_{ij}) 和 (x'_{ij}, y'_{ij}) 分别表示点 p_{ij} 在网格变形前和变形后的坐标,这里, $x_{ij}=i, y_{ij}=j$.因为只考虑水平方向的网格变形,我们规定每个点的 y 坐标在变形中不发生变化,即 $y_{ij} = y'_{ij}$.

由于曲边梯形网格的边缘中包含曲线段,网格变形后的形状无法直接根据重定位的顶点来确定.因此,我们

对网格变形中曲边上的各点坐标变化进行了约束.对于网格右侧边缘上的任意一点 p_{Rj} ,我们找到与其 y 坐标相同且位于右侧边缘两个顶点连线上的点 $p_{R'j}$,并规定点 p_{Rj} 和 $p_{R'j}$ 水平距离在网格变形中必须与点 p_{RT} 和 p_{RB} 之间的水平距离成正比,即

$$\frac{x_{R'j} - x_{Rj}}{x_{RB} - x_{RT}} = \frac{x'_{R'j} - x'_{Rj}}{x'_{RB} - x'_{RT}} \quad (1)$$

该规定保证了网格右侧边缘上各点的变化在网格变形中可以保持较好的一致性.当 $x_{RT}=x_{RB}$ 时,我们规定 $x'_{Rj} = x'_{RT} = x'_{RB}$.由于任意点在网格变形前后 y 坐标不变,且点 $p_{R'j}$ 位于点 p_{RT} 和 p_{RB} 之间的连线上,公式(1)可以改写为

$$x'_{Rj} = \frac{x_{Rj} - x_{RB}}{x_{RT} - x_{RB}} x'_{RT} + \frac{x_{RT} - x_{Rj}}{x_{RT} - x_{RB}} x'_{RB} \quad (2)$$

由于 x_{Rj}, x_{RT}, x_{RB} 均为已知量,由公式(2)可知,点 p_{Rj} 在网格变形后的 x 坐标 x'_{Rj} 为 x'_{RT} 和 x'_{RB} 的线性组合.同样,我们对于位于网格左侧边缘上的任意一点 p_{Lj} 做类似的约束,其 x 坐标 x'_{Lj} 在网格变形后也可以表示为 x'_{LT} 和 x'_{LB} 的线性组合.根据上述约束,变形后网格的形状可以根据原始网格的形状和网格顶点的新坐标来确定.

4.2 带约束的网格变形优化

在单个网格变形约束的基础上,我们根据区域关系图和能量模板来建立优化方程,并通过优化各网格的面积变化和形状变化来分别实现突出重要物体和消除视觉扭曲的目的.

对于网格的面积变化,由于单个网格变形后的面积 S' 可看作其所包含的像素数量之和,即

$$S' = \sum_{j=y_{RB}}^{y_{RT}} (x'_{Rj} - x'_{Lj}).$$

将公式(2)代入, S' 则可表示为 $x'_{LT}, x'_{LB}, x'_{RT}, x'_{RB}$ 的线性组合

$$S' = \alpha_{LT} x'_{LT} + \alpha_{LB} x'_{LB} + \alpha_{RT} x'_{RT} + \alpha_{RB} x'_{RB} \quad (3)$$

其中, $\alpha_{LT}, \alpha_{LB}, \alpha_{RT}, \alpha_{RB}$ 为根据公式(2)生成的系数.

我们通过最小化网格面积变化的加权平方和实现突出重要物体的目的,相应的惩罚函数被定义为

$$f_s = \sum_k e_k^i (S'_k - S_k)^2 \quad (4)$$

其中, S_k 和 S'_k 分别表示第 k 个网格在变形前后的面积, e_k^i 表示第 k 个网格中像素重要度能量的平均值.通过最小化函数 f_s ,重要度能量越高的网格将在变形中发生越小的面积变化,从而达到被突出的目的.

除了考虑网格的面积变化外,我们通过优化网格的形状变化来减少目标图像中的视觉扭曲.假设单个网格内部同一行的像素在网格变形过程中采用均匀采样,则网格内任意一点 p_{ij} 在网格变形后的 x 坐标 x'_{ij} 为

$$x'_{ij} = x'_{Lj} + \frac{x_{ij} - x_{Lj}}{x_{Rj} - x_{Lj}} (x'_{Rj} - x'_{Lj}) \quad (5)$$

由于 x_{ij}, x_{RT}, x_{RB} 均为已知量,结合公式(2)可知, x'_{ij} 可以表示为网格变形后各个顶点 x 坐标的线性组合.

我们用网格内每个像素与正上方像素在网格变形后水平方向距离的加权平方和来衡量网格形状变化程度,并通过考察网格左/右边缘上每个点与其上方在边缘上点的水平距离变化来避免类似于长宽比不固定的均匀缩放所产生的扭曲,网格形状的变化程度用公式表示如下:

$$D = \sum_{j=y_{RB}}^{y_{RT}} \frac{1}{x_{Rj} - x_{Lj}} \sum_{i=x_{Lj}}^{x_{Rj}} e_{ij}^{HS} e_{i(j+1)}^{HS} (x'_{ij} - x'_{i(j+1)})^2 + \sum_{j=y_{RB}}^{y_{RT}} \frac{1}{2} (e_{Lj}^{HS} e_{L(j+1)}^{HS} ((x'_{Lj} - x'_{L(j+1)}) - (x_{Lj} - x_{L(j+1)}))^2 + e_{Rj}^{HS} e_{R(j+1)}^{HS} ((x'_{Rj} - x'_{R(j+1)}) - (x_{Rj} - x_{R(j+1)}))^2) \quad (6)$$

其中, e_{ij}^{HS} 和 $e_{i(j+1)}^{HS}$ 分别表示用于水平方向缩放的敏感度模板在点 p_{ij} 和 $p_{i(j+1)}$ 处的能量值.将公式(5)代入,公式(6)可改写为

$$D = \sum_{j=y_{RT}}^{y_{RT}} \left(\frac{\sum_{i=x_{Lj}}^{x_{Rj}} e_{ij}^{HS} e_{i(j+1)}^{HS}}{x_{Rj} - x_{Lj}} + \frac{e_{Lj}^{HS} e_{L(j+1)}^{HS} + e_{Rj}^{HS} e_{R(j+1)}^{HS}}{2} \right) (\beta_{LT} x'_{LT} + \beta_{LB} x'_{LB} + \beta_{RT} x'_{RT} + \beta_{RB} x'_{RB})^2 \quad (7)$$

其中, $\beta_{LT}, \beta_{LB}, \beta_{RT}, \beta_{RB}$ 为根据公式(2)、公式(5)、公式(6)生成的系数。

我们将各个网格的形状变化程度之和作为控制视觉扭曲的惩罚函数:

$$f_D = \sum_k D_k \quad (8)$$

其中, D_k 表示第 k 个网格的形状变化程度. 通过最小化函数 f_D , 敏感度能量越高的网格在变形中发生越小的形状变化, 从而达到消除明显视觉扭曲的目的。

我们将网格面积变化和形状变化的惩罚函数合并到目标函数中, 对突出重要物体和消除明显视觉扭曲两个目标同时进行优化. 这样可以避免将某个目标作为约束条件而非优化对象时因约束过松或过紧而造成的效果欠佳问题. 优化目标函数定义为

$$\min w_S f_S + w_D f_D \quad (9)$$

其中, w_S 和 w_D 为非负系数, 且满足 $w_S + w_D = 1$. w_S 越大时, 目标图像中的重要物体越突出, 但越容易引起明显的视觉扭曲; w_D 越大时, 目标图像中的视觉扭曲越不明显, 但重要物体会越不突出. 在本文实验中, $w_S = w_D = 0.5$.

由公式(3)、公式(4)、公式(7)、公式(8)可知, 该优化问题是一个关于网格变形后各网格顶点 x 坐标的二次规划问题. 在建立优化方程时, 我们仅将位于图像内部关键点的横坐标作为变量引入(图 3(c)中的圆形标记); 而对位于图像内的辅助点(三角标记), 它们在网格变形后的横坐标根据公式(2)使用相应的关键点表示; 对位于图像边缘上的关键点和辅助点(星形标记和方形标记), 我们规定它们在网格变形后必须落在目标图像相应的边缘上. 上述处理可以在网格变形中有效地保持边缘的整体形态, 并且进一步减少了优化方程中的变量数目, 降低了问题求解的计算量. 例如, 图 3(c)中本文方法所需重定位的网格顶点数量仅为 14 个, 而图 3(e)和图 3(f)中需要重定位的网格顶点数量分别为 108 个和 960 个. 图像内关键点的判别可以通过扫描区域关系图来实现. 如果某一层中两个相邻结点的父结点(或子结点)均与自身编号不相同, 且它们的上底边(或下底边)共用顶点位于图像内, 则该共用的顶点为位于图像内的关键点. 位于最上层(或最下层)的结点可以看作是父结点(或子结点)与自身编号不相同的特殊情形. 除图像的 4 个顶点外, 位于图像上/下边缘的网格顶点均应看作是图像内的关键点。

根据区域关系图, 我们在优化方程的约束条件中规定: 每层第 1 个结点的左边缘顶点在网格变形后 x 坐标为 1, 最后一个结点的右边缘顶点在网格变形后 x 坐标为 W (W 为目标图像的宽度). 此外, 对位于图像内辅助点形成的网格顶点, 我们根据单个网格的变形约束将其与对应网格顶点的关系作为约束条件加入. 这种辅助点类的网格顶点可以通过在区域关系图中查找相同编号的父结点和子结点来发现. 如果它们左侧的父子结点编号相同, 则辅助点位于对应网格的右边缘上; 如果右侧的父子结点编号相同, 则辅助点位于对应网格的左边缘上. 为了避免产生伪解, 我们还在优化方程中增加了 3 类约束条件: (1) 位于图像中同一行的网格顶点的左右相对位置在变形前后保持不变, 即如果 $x_{i,j} < x_{i_2,j}$, 则 $x'_{i,j} \leq x'_{i_2,j}$; (2) 同一网格左/右边缘两个顶点的左右相对位置在网格变形前后保持不变, 即如果 $x_{*T} \neq x_{*B}$, 则 $(x_{*T} - x_{*B})(x'_{*T} - x'_{*B}) \geq 0$. 这里, $*$ 表示 L 或 R ; (3) 网格上/下底边在网格收缩时不增长, 即 $x'_{Rj} - x'_{Lj} \leq x_{Rj} - x_{Lj}$; 在网格拉伸时不缩短, 即 $x'_{Rj} - x'_{Lj} \geq x_{Rj} - x_{Lj}$.

根据网格变形后网格顶点的新坐标, 我们计算出原始图像中每个像素在目标图像中的位置, 再通过插值生成目标图像. 对于区域关系图中一个结点拥有多个父结点或子结点的情形, 我们采用迭代平滑的方法对位于网格交界处像素的新坐标进行修正, 以缓解直接均匀插值出现“割裂”现象。

4.3 适配顺序和变形方向

由于水平方向适配和竖直方向适配被分开处理, 显示适配过程中存在“水平-竖直”和“竖直-水平”两种可能的适配顺序. 图 5 展示了对相同原始图像按这两种适配顺序分别生成相同尺寸目标图像的示例. 实验表明, 按照不同适配顺序生成的目标图像均能够有效地保留原始图像的信息, 且具有较高的相似性. 这表明本文方法对适配顺序具有较好的鲁棒性。

此外, 由于在某一方向上收缩图像和在另一方向上拉伸图像可以得到长宽比相同但效果迥异的目标图像,

显示适配中的变形方向也是需要考虑的问题^[13].在本文中采用结合预缩放(preliminary scaling)的策略来确定变形的方向.我们首先根据每个像素的重要度能量计算每行的“理想长度” $L_i = \sum_{j=1}^W (e'_{ij})^\gamma$,其中: W 为原始图像的宽度; γ 表示突出重要物体的程度,可以通过手工指定或自动计算(本文实验中 γ 为优化方程目标函数中 w_s 系数的倍数 $\gamma=Nw_s, N=100$).接着,对于图像中所有行,我们找出各行中理想长度的最大值 $L_{Hmax} = \max_i L_i$.同样,我们求出每列的理想长度 L_j ,并求出各列中理想长度的最大值 $L_{Vmax} = \max_j L_j$.根据 L_{Hmax} 和 L_{Vmax} 计算出预缩放的比例 $\rho = \min\left\{\frac{L_{Hmax}}{W'}, \frac{L_{Vmax}}{H'}\right\}$.其中, W' 和 H' 分别表示目标图像的宽和高.我们将原始图像在水平和垂直方向上都均匀缩放到原来的 $1/\rho$,再根据目标屏幕的尺寸对缩放后的图像在水平和垂直方向上进行收缩或拉伸.

结合预缩放不仅可以较好地解决变形方向选择的问题,还可以使得缩放后的图像中重要物体的大小满足目标图像尺寸,避免了重要物体在优化求解中的变形,因而能够较好地保持其形状.与多能量模板约束策略一样,结合预缩放的策略也可推广应用到现有的其他显示适配方法中.

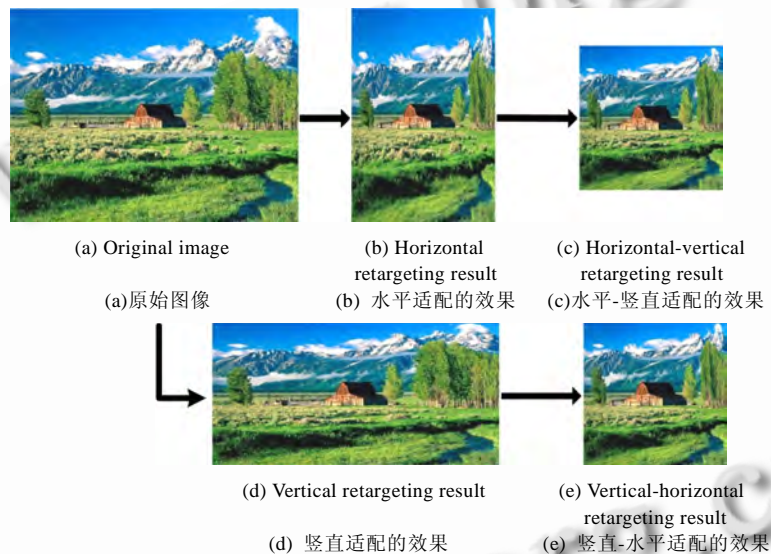


Fig.5 Results with different retargeting orders

图 5 根据不同适配顺序生成的效果

5 实验结果与分析

为了验证本文方法的效果,我们实现了该方法并在多个方面进行了测试.实验中使用的机器CPU为Intel P4 3.0GHz,内存为2GB,编程语言为Matlab.

5.1 实验结果

我们在包含30幅图像的测试集上对本文方法进行了实验.测试集中包含的图像具有不同内容和风格,分辨率从 $800 \times 600 \sim 960 \times 600$ 不等.我们规定目标图像的分辨率为 300×300 .

图6展示了由同一原始图像生成不同长宽比目标图像的示例,目标图像包括了2.4:1,16:9,4:3,1:1,3:4等常见的长宽比.实验结果表明,本文方法对于不同长宽比的目标屏幕需求均能够取得较好的适配效果.

图7展示了本文方法与现有典型方法的比较示例.我们选择了7种代表性方法进行比较,包括裁剪方法^[1]、基于分割的重构方法^[6]、抽丝方法^[7]、非均匀重定位方法^[9]、鱼眼变形方法^[11]、基于网格参数化的方法^[12]和基于缩放-拉伸的方法^[13].可以发现:裁剪方法无法在目标图像中保留所有的重要物体;基于分割的重构方法无法很好地处理位于边缘的物体,从而造成目标图像中物体空间位置关系和尺寸比例的失真;抽丝方法在物体内

部梯度较低时会破坏物体的结构;非均匀重定位方法会造成“自相交”现象,使得部分重要内容丢失;基于鱼眼变形的的方法需要手工选择变形模板,并且在没有关注的区域会形成明显扭曲;基于网格参数化的方法在图像边缘处扭曲了未能成功检测出的物体;基于缩放-拉伸的方法忽略了突出重要物体,并因物体梯度较小而造成物体结构的破坏.与上述方法相比,本文方法突出了重要物体并且没有明显的视觉扭曲,具有更好的显示适配效果.

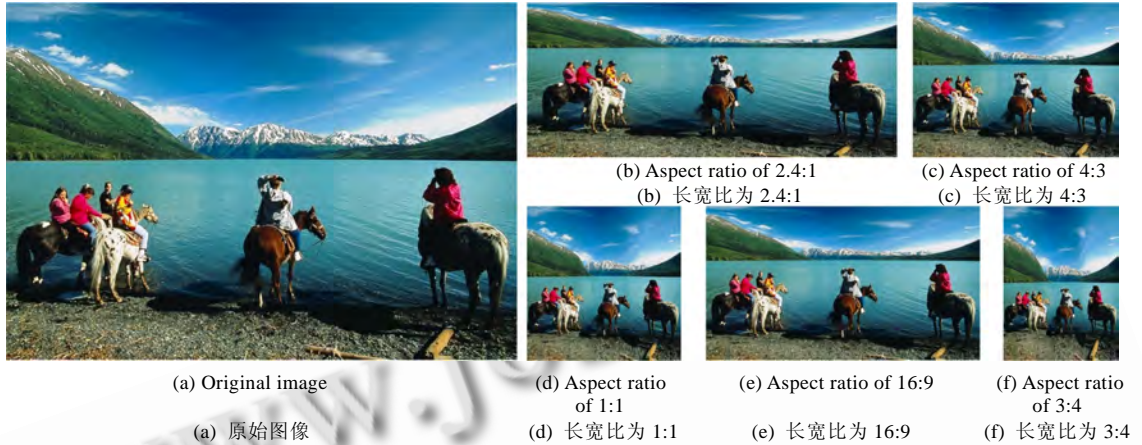


Fig. 6 Results with different aspect ratios

图 6 不同长宽比的适配效果

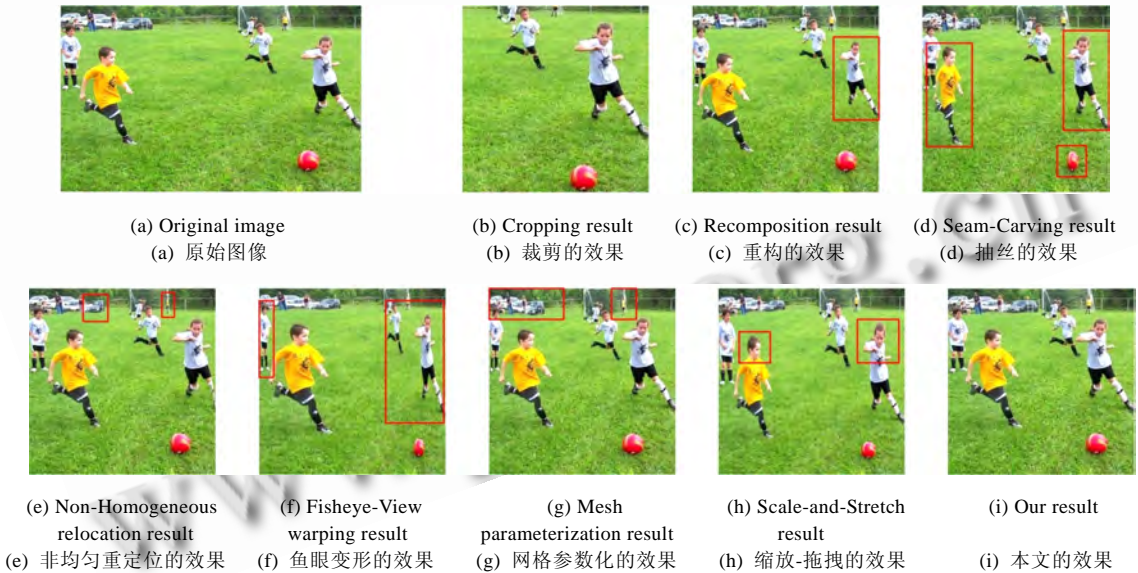


Fig.7 Example of comparison with the existing image retargeting methods

图 7 与现有图像显示适配方法比较的示例

5.2 用户调查

由于目前没有针对图像显示适配效果的自动评估机制,我们采用用户调查的评估方式来进一步检验本文方法的有效性.我们邀请了 21 名评估者参加用户调查,其中男性 12 人、女性 9 人.评估者的年龄在 20 岁~45 岁之间,职业包括学生、教师、公司职员、政府工作人员等.据我们所知,所有用户调查均在评估者对本文研究内容不知情的状态下进行.用户调查的内容分为两部分:第 1 部分是对本文方法效果的直接评估,第 2 部分是与现

有其他显示适配方法效果的比较。

在用户调查的第 1 部分,评估者通过比照原始图像来判断根据本文方法生成的目标图像是否包含了原始图像的基本信息(包括重要物体的细节),以及在视觉质量上是否令人满意(没有明显的视觉扭曲)。每个评估者分别独立对 30 个结果进行评价。对前一个问题表示肯定的比例为 95.7%,对后一个问题表示肯定的比例为 91.9%。

在用户调查的第 2 部分,评估者通过对比本文方法和现有其他方法的适配效果来评价本文方法的优劣。对于测试集中每一幅原始图像,我们将由本文方法生成的目标图像分别与上述 7 种代表性的方法所生成的目标图像一一组对。每对目标图像由随机选择的 3 名评估者在不知道相应生成方法的情况下评定某个目标图像效果更好或两者效果相似。对于同一原始图像产生的 7 个目标图像对,每个评估者能够且仅能够参与其中之一的评判。3 名评估者意见中的多数意见被认定为最终评价;如果 3 名评估者意见均不同,则认为两种方法效果相似。表 1 展示了用户调查的结果。结果表明,本文方法的效果优于现有的绝大多数方法,且与基于网格参数化的方法相近。但由于本文方法和基于网格参数化的方法都是采用带约束的二次规划,且在通常情况下本文方法所使用的变量个数(测试集上平均每幅图像对应的变量数为 23.4 个)会远远小于基于网格参数化的方法(约 120 个^[12]),因此本文方法具有更高的效率。此外,本文方法不需要根据重要度判断网格是否包含物体,且不会将视觉扭曲集中在判定为包含重要物体的网格周围,因此对重要度能量的计算结果会更加鲁棒。

Table 1 Result of user study in comparison with the existing typical methods

表 1 与现有典型方法比较的用户调查结果

	Better	Similar	Worse
Cropping	25	2	3
Segmentation based recombination	22	6	2
Seam carving	18	9	3
Non-Homogeneous relocation	20	8	2
Fisheye-View warping	24	5	1
Scale-and-Stretch	15	13	2
Mesh parametrization	9	17	4

由于能量模板的计算和区域划分均可以在预处理阶段完成,本文方法的效率主要取决于区域关系图的构建和网格顶点重定位的优化求解。在本文的实验环境下,测试集上平均每幅图像的处理时间为 2.37s。通过使用高效编程语言和优化求解过程,本文方法的效率有望得到大幅度的提高。

5.3 本文方法的局限性

在实验中,我们也发现了本文方法的一些缺陷。例如,本文方法力图在突出重要物体和消除视觉扭曲之间取得最佳的平衡。然而对于某些情形,只考虑其中某一个因素却会取得更好的效果。在图 8(a)、图 8(b)展示的布达拉宫的图像中,由于布达拉宫的知名度,保证建筑主体不发生任何形变要比同时保留显著的附属建筑更受评估者关注。此时,使用裁剪方法(如图 8(b)所示)可以取得比本文方法(如图 8(a)所示)更好的效果。此外,除了重要物体的突出和视觉扭曲的消除外,图像显示适配的效果中还受其他因素的影响。图 8(c)、图 8(d)展示的是绿叶背景下的花丛。均匀缩放的结果(如图 8(d)所示)虽然没有突出重要物体(花丛)也引起了明显的视觉扭曲,但它保持了图像的构图信息(花丛与绿叶的比例),因而比本文方法的效果(如图 8(c)所示)更受用户欢迎。

6 结束语

本文提出一种新颖的基于区域关系图的图像显示适配方法。该方法将图像用区域关系图表示,并将图像显示适配问题规约为关于关键网格顶点重定位的二次规划问题进行求解。本文方法率先采用了多能量模板约束策略,通过不同的能量模板来约束图像显示适配效果的不同方面。本文还提出了曲边梯形网格表示及区域关系图的构建方法,可以更好地描述图像结构和减少计算量。在构建优化方程的过程中,本文方法将突出重要物体和消除明显视觉扭曲统一在目标函数中,根据区域关系图建立相应的约束条件,并结合预缩放来改进效果。实验表明,本文方法适用于不同风格图像和目标屏幕长宽比的显示适配,并且能够取得比现有其他方法更好的效果。



Fig.8 Examples of our limitation

图8 本文方法局限性的示例

在今后的工作中,我们将针对本文方法中存在的缺陷进行改进,例如与裁剪方法和均匀缩放方法的结合等.我们还计划将本文方法推广到视频领域,实现面向视频的显示适配技术也是未来的研究方向.

致谢 在此,我们向对本文工作给予宝贵意见的审稿专家表示衷心感谢.此外,感谢台湾成功大学的王昱舜博士和南京大学的郭延文副教授提供了他们各自方法的实验结果.

References:

- [1] Chen LQ, Xie X, Fan X, Ma WY, Zhang HJ, Zhou HQ. A visual attention model for adapting images on small displays. *Multimedia Systems*, 2003,9(4):353–364. [doi: 10.1007/s00530-003-0105-4]
- [2] Wang Y, Li HQ, Liu ZK, Chen CW. Attention information based spatial adaptation framework for browsing videos via mobile devices. *Advances in Multimedia Information Processing*, 2006,4621:788–797. [doi: 10.1007/11922162_90]
- [3] Deselaers T, Dreuw P, Ney H. Pan, zoom, scan—time-coherent, trained automatic video cropping. In: Ahuja N, Shapiro L, eds. *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition*. Anchorage: IEEE Press, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587729]
- [4] Xie X, Liu H, Ma WY, Zhang HJ. Browsing large pictures under limited display sizes. *IEEE Trans. on Multimedia*, 2006,8(4): 707–715. [doi: 10.1109/TMM.2006.876294]
- [5] Jiang SQ, Liu HY, Zhao Z, Huang QM, Gao W. Generating video sequence from photo image for mobile screens by content analysis. In: Gao W, ed. *Proc. of the Int'l Conf. on Multimedia and Expo*. Beijing: IEEE Press, 2007. 1475–1478. [doi: 10.1109/ICME.2007.4284940]
- [6] Setlur V, Lechner T, Nienhaus M, Gooch B. Retargeting images and video for preserving information saliency. *IEEE Computer Graphics and Applications*, 2007,27(5):80–88. [doi: 10.1109/MCG.2007.133]
- [7] Rubinstein M, Shamir A, Avidan S. Improved seam carving for video retargeting. *ACM Trans. on Graphics*, 2008,27(3):16. [doi: 10.1145/1360612.1360615]
- [8] Huang H, Fu TN, Rosin PL, Qi C. Real-Time content-aware image resizing. *Science in China Series F: Information Sciences*, 2009, 52(2):172–182. [doi: 10.1007/s11432-009-0041-9]
- [9] Wolf L, Guttman M, Cohen-Or D. Non-Homogeneous content-driven video-retargeting. In: Davis L, Bouthemy P, Ikeuchi K, eds. *Proc. of the Int'l Conf. on Computer Vision*. Rio de Janeiro: IEEE Press, 2007. 1–6. [doi: 10.1109/ICCV.2007.4409010]
- [10] Ren TW, Liu Y, Wu GS. Image retargeting based on global energy optimization. In: Lin CY, Cox I, eds. *Proc. of the Int'l Conf. on Multimedia and Expo*. New York: IEEE Press, 2009. 406–409. [doi: 10.1109/ICME.2009.5202520]
- [11] Liu F, Gleicher M. Automatic image retargeting with fisheye-view warping. In: Czerwinski M, Baudisch P, eds. *Proc. of the Symp. on User Interface Software and Technology*. Seattle: ACM Press, 2005. 153–162. [doi: 10.1145/1095034.1095061]
- [12] Shi J, Guo YW, Du ZL, Zhang FY, Peng QS. A mesh parameterization-based image retargeting method. *Journal of Software*, 2008, 19:19–30 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/s19.htm>

- [13] Wang YS, Tai CL, Sorkine O, Lee TY. Optimized scale-and-stretch for image resizing. *ACM Trans. on Graphics*, 2008,27(5):1–10. [doi: 10.1145/1409060.1409071]
- [14] Ren TW, Guo YW, Wu GS, Zhang FY. Constrained sampling for image retargeting. In: Ostermann J, ed. *Proc. of the Int'l Conf. on Multimedia and Expo*. Hannover: IEEE Press, 2008. 1397–1400. [doi: 10.1109/ICME.2008.4607705]
- [15] Ren TW, Liu Y, Wu GS. Image retargeting using multi-map constrained region warping. In: Gao W, Rui Y, Hanjalic A, eds. *Proc. of the Int'l Conf. on Multimedia*. Beijing: ACM Press, 2009. 853–856. [doi: 10.1145/1631272.1631431]
- [16] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998,20(11):1254–1259. [doi: 10.1109/34.730558]
- [17] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Kasturi R, Medioni G, eds. *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition*. Kauai: IEEE Press, 2001. 511–518. [doi: 10.1109/CVPR.2001.990517]
- [18] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002,24(5):603–619. [doi: 10.1109/34.1000236]
- [19] Bober M. MPEG-7 visual shape descriptors. *IEEE Trans. on Circuits and Systems for Video Technology*, 2001,11(6):716–719. [doi: 10.1109/76.927426]

附中文参考文献:

- [12] 时健,郭延文,杜振龙,张福炎,彭群生.一种基于网格参数化的图像适应方法.软件学报,2008,19:19–30. <http://www.jos.org.cn/1000-9825/19/s19.htm>



任桐炜(1981—),男,江苏南通人,博士生,主要研究领域为图像处理和检索.



武港山(1967—),男,博士,教授,博士生导师,主要研究领域为多媒体信息处理,信息检索.



刘焱(1974—),女,博士,助理教授,博士生导师,主要研究领域为多媒体分析.