

# 基于模糊估计融合显著性检测的自动抠图算法

裴晓康, 王贵锦, 林行刚

(清华大学 电子工程系, 北京 100084)

**摘要:** 提出了一种新的自动抠图算法框架。首先,估计输入图像各个区域的模糊程度;其次,对图像进行显著性的计算;然后融合模糊度和显著性信息,大致分割出前景和背景,从而自动生成标注前景、背景、未知区域的三色图;最后,采用基于采样的抠图算法就可以准确地完成前景目标物体的自动抠图。该算法无须人工辅助或附加信息。在标准数据集和实拍图像上的实验结果证明了该算法的准确性和实用性,可广泛应用于图像和视频的编辑合成。

**关键词:** 自动抠图; 自然背景; 模糊估计; 显著性检测; 三色图生成

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-3695(2012)10-3945-03

**doi:**10.3969/j.issn.1001-3695.2012.10.092

## Automatic matting based on fusion of blur estimation and saliency detection

PEI Xiao-kang, WANG Gui-jin, LIN Xing-gang

(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)

**Abstract:** This paper proposed a new framework for automatic matting of natural images. First, estimated the blurring level of each part of the input image according to the different focusing status in the foreground and in the background. Second, it calculated the saliency for the image and detected the salient regions. Then, with the combined information of blur and saliency, the foreground region and background region could be roughly identified and separated, and the Trimap which denoted the division of the foreground, background and unknown region could be generated automatically. Last, accurate object matting could be accomplished by a sample-based matting algorithm on the basis of the Trimap. This algorithm did not require any user assistance or other additional information. Experiments both on standard data sets and actual captured images well demonstrate the accuracy and effectiveness of this algorithm, which can be widely used in image/video editing and composition.

**Key words:** automatic image matting; natural background; blur estimation; saliency detection; Trimap generating

## 0 引言

自然背景下的数字图像抠图(digital matting)是指在图像中将前景目标物体精确提取的过程。它是图像处理与编辑中的一项关键技术,可广泛应用于视频合成、电影特效以及游戏娱乐等领域。与图像分割技术不同的是,抠图是一种“软分割”的技术,它需要将图像中每个像素精确地表示为前、背景的按比例合成。其数学模型表示为:对自然图像的任意像素 $I_i$ ,可以表示为背景像素 $B_i$ 和前景像素 $F_i$ 的线性合成,如式(1)所示。

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \quad (1)$$

其中: $\alpha_i$ 就是所要求的每个像素的合成比,反映的是该像素的不透明度。由于前景像素值 $F_i$ 和背景像素值 $B_i$ 也是未知的,所以抠图问题是一个不能完全求得解析解的病态问题。而为了解决这个问题,一般需要提供一个人工指定的标明数字图像上的确定前景、背景和未知的区域,并分别用三种颜色表示,称为三色图(Trimap)。有了Trimap就可以提供估计前背景的依据,减少了未知量的计算。在此基础上研究者提出了很多依赖人工标定Trimap的半自动的抠图算法,Wang等人<sup>[1]</sup>对这些算法进行了很好的总结。

人工标定图像的Trimap毕竟是一件费时费力的工作,尤其是对视频抠图与合成。因此,一些自动抠图方法也相继被研究者提出。McGuire等人<sup>[2]</sup>在2005年首先提出了一种针对动态场景的自动抠图方法,称为散焦抠图(defocus matting)。散焦抠图利用分光镜搭建了一个由三个摄像头传感器组成的系统,对同一场景采集聚焦前景、聚焦背景以及全景深聚焦这样三幅图片,然后通过聚焦面得到场景物体的相对深度,从而自动计算出粗略的Trimap,进而转换为全局约束优化问题用梯度下降法求出未知区域像素的 $\alpha$ 值和真实的前背景颜色。Sun等人<sup>[3]</sup>提出了闪光抠图(flash matting)的方法。该方法的思路是如果背景物体离得足够远,那么场景的加闪光照片与不加闪光照片之间的差别主要集中在前景物体区域。因此闪光抠图先采集同一场景加闪光和不加闪光两幅图片,然后通过两幅图像的差估计每个像素属于前景的似然概率,在联合贝叶斯模型的框架下迭代求最大似然解。此外,还有利用照相机阵列<sup>[4]</sup>和利用深度传感器<sup>[5]</sup>的抠图方法被提出。这些自动抠图方法虽然避免了人工标定的步骤,但却都需要提供额外的设备(如多传感器、闪光灯等),利用这些设备获取单张彩色图像之外的信息来估计前背景的区分,难以用于实际自然背景下的图像抠图。

为了克服以上这些方法的弱点,本文提出了一种无须依赖外部设备或信息,只以单幅彩色图像作为输入的自动抠图算法

收稿日期: 2012-03-20; 修回日期: 2012-04-25

作者简介:裴晓康(1988-),男,江苏东台人,硕士,主要研究方向为图像处理、视觉计算(pxk09@mails.tsinghua.edu.cn);王贵锦(1976-),男,副教授,博导,博士,主要研究方向为图像处理与模式识别算法;林行刚(1947-),男,教授,博导,博士,主要研究方向为视频分析。

框架。本算法的主要思路是从输入图像不同区域的模糊程度和显著性度量值两个方面来判断该区域是否属于前景目标,从而自动生成粗略的 Trimap,然后再采用一般的半自动抠图算法计算出精确的抠图结果。与散焦抠图<sup>[2]</sup>相似,本文提出的框架也要求摄像机聚焦前景目标物体,并且前背景分开一定的距离,使得背景图像产生散焦虚化。但不同的是,本文框架融合散焦估计和显著性检测来较为鲁棒地直接估计前背景区域,而不需要额外的传感器采集前景虚化和全景深的两幅图像,大大减少了对系统的要求。事实上,在视频合成中抠图的作用往往是准确提取清晰的前景图像合成到其他背景中,对原背景是否聚焦并无要求,所以本文的方法在简化系统、自动实现的同时又不失广泛适用性。

## 1 算法的框架流程

基于模糊估计融合显著性检测的自动抠图算法通过挖掘输入图像的多种信息来估计反映场景远近的相对深度,计算前背景的粗略分割结果,提供 Trimap 从而完成自动抠图。本文算法的流程如图 1 所示,主要分为四个步骤:a)计算图像每个像素的散焦模糊程度;b)计算图像每个区域的显著性;c)融合模糊度和显著性,得到 Trimap;d)引入基于采样的抠图算法进行抠图。

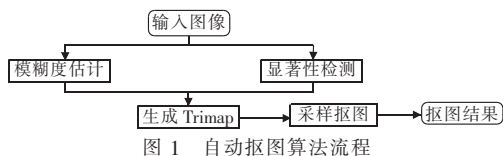


图 1 自动抠图算法流程图

## 2 算法的详细介绍

### 2.1 散焦模糊度估计

根据相机的成像原理,对于镜头尺寸较大的相机,它的景深范围较小,随深度变化的散焦模糊效应也较明显。因此散焦程度在一定程度上反映了对应物体的深度。在计算机视觉中,从散焦恢复深度(depth from defocus)就是模糊估计的一项重要应用,但是它往往需要两张或更多张不同聚焦深度的图片去完整恢复出场景的深度,而本文的目标是估计聚焦前景的图片不同位置的散焦模糊程度,从而得到一张反映该程度的模糊图(blur map)。

Bae 等人<sup>[6]</sup>提出了一种散焦估计算法用于背景虚化增强。由于该算法计算复杂度较高,本文提出了一种改进的散焦估计的算法。考虑到聚焦的前景物体往往包含更多的细节和边缘,本文的模糊估计算法从图像的边缘出发,估计它们的模糊程度,然后再统计各个区域平均的边缘强度,将边缘所在区域求得的模糊程度扩展传播到其他区域。实现细节如下:

a)对图像进行预处理,将彩色图像转换为灰度图像,并进行小尺度的平滑滤波,以降低高频图像噪声的影响。散焦对物体纹理和边缘的模糊效应可以近似为高斯模糊,其模糊核为

$$g(x, y, \sigma_b) = \frac{1}{2\pi\sigma_b^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_b^2}\right) \quad (2)$$

其中: $\sigma_b$  是模糊核的尺度,代表模糊的程度。

b)对预处理后的图像进行边缘检测,得到反映边缘位置和强度的边缘图,然后再进行大尺度的高斯滤波平滑,就得到了初始的模糊程度图。

c)由边缘处的模糊度扩散估计全图的模糊度。本文对初始的模糊程度图设计了一系列窗口尺寸由小到大的平均滤波

器,用于检测每个位置最小的均值和最大的均值,以这两个值按固定权重的平均值作为该位置的模糊度。

d)对模糊图进行增强优化,对包含边缘细节较多的区域减少其模糊度值以降低多次平滑的影响,完成最终模糊程度图的估计。

本文算法中使用 Canny 边缘检测算子,大尺度的高斯平滑核半径取为 6。图 2 是计算过程的一组样例示意,包含了中间步骤的结果。



图 2 散焦模糊度估计的计算过程

### 2.2 显著性区域检测

虽然散焦模糊程度可以很好地区别前景和背景,但是它依赖于边缘检测,受场景物体纹理影响比较大,因此需要进一步的计算信息来更加鲁棒地识别前景区域。视觉显著性检测(saliency detection)是近年来计算机视觉领域的一个热门课题,它的目标是在单幅图像中检测各个区域的显著性也就是对观察者的吸引程度,一个重要的应用方向就是对感兴趣目标物体的图像分割。针对颜色平滑且不包含较多纹理的前景物体,本文引入了区域显著性检测进行标志。

关于显著性检测算法,研究人员基于生物学原理和视觉计算等提出了很多不同算法。其中 Cheng 等人<sup>[7]</sup>提出的基于全局对比度的算法简单高效且具有很好的准确性,因此本文算法框架引入文献[7]的算法作为显著性区域检测模块。算法基于图像的颜色统计特征,采用直方图对比度来定义每个像素的显著性,即

$$S(I_k) = \sum_{I_l \in I} D(I_k, I_l) \quad (3)$$

其中: $I_k$  是图像  $I$  中的任意像素, $D(I_k, I_l)$  为两像素在 Lab 色彩空间中的距离度量。把具有相同颜色值  $c_j$  的像素放到一起,式(3)可以重写为某颜色值的显著性:

$$S(c_l) = \sum_{j=1}^n f_j D(c_l, c_j) \quad (4)$$

其中: $n$  是图像中出现的颜色总数, $f_j$  是颜色  $c_j$  在图像中出现的概率,表现为颜色直方图上的某个分量。

具体实现中引入色彩量化减小直方图的尺寸,并对显著性函数曲线进行颜色空间的加权平滑。为了提高显著性检测的准确性,先将图像分割至若干块,对每一块建立颜色直方图,然后以整块为单位求它跟其他区域颜色的对比度,以此作为该区域的显著性。这样引入了图像的空间位置信息后,使得结果更为准确。本文实验中,参数设置与文献[7]一致,颜色数量化到约 85,区域空间权重强度取为 0.4。

一组典型的检测结果如图 3 所示,显著性检测得到的显著分布图可以较好地地区分前/背景,尤其在深度物体的边缘处更加准确。



图 3 显著性区域检测的效果示意图

### 2.3 Trimap 计算

在 2.1 和 2.2 节中已经分别计算出了模糊程度图和显著性图。将两者融合到一起就可以提供较为鲁棒的前/背景区分度的概率分布。

设任意像素  $p$  的模糊度为  $C_{\text{blur}}(p)$ , 显著性为  $C_{\text{saliency}}(p)$ , 设计求解该像素属于前景的概率  $C_{\text{foreground}}(p)$  的融合式为

$$C_{\text{foreground}}(p) = \frac{1}{2}(\rho(C_{\text{blur}}(p), \lambda_b) + \rho(C_{\text{saliency}}(p), \lambda_s)) \quad (5)$$

其中:  $\rho(c, \lambda)$  为非线性变换函数, 写为

$$\rho(c, \lambda) = 1 - \exp(-\frac{c}{\lambda}) \quad (6)$$

参数  $\lambda_b$  与  $\lambda_s$  用于控制模糊度和显著性的融合比例, 并抑制噪声点的影响。

对前景概率图进行阈值分割, 得到粗略估计的前景区域  $\Omega_F$  和背景区域  $\Omega_B$ 。再对得到的二值图进行一次闭运算消除前景区域内部的一些空洞。然后分别对前景区域和背景区域进行宽度为  $b$  的腐蚀操作, 两次操作的差即为未知区域  $\Omega_U$ , 如式(7)所示。这样就生成了标注前景、背景和未知区域的 Trimap。

$$\Omega_U = \overline{\text{erode}(\Omega_F, b) \cup \text{erode}(\Omega_B, b)} \quad (7)$$

### 2.4 抠图

通过前面的步骤已经自动计算出了 Trimap, 接下来基于 Trimap 的一般抠图算法都适合本文的框架。考虑到只从待抠图的图像本身出发计算的 Trimap 在精确性方面不如人工标定的 Trimap, 因此本文选择了一种基于采样和全局最优精化的鲁棒抠图算法(robust matting)<sup>[8]</sup>作为最终的抠图模块, 该抠图算法准确度高且对 Trimap 依赖度较小。

鲁棒抠图算法<sup>[8]</sup>的大致步骤为: a) 对未知区域的像素, 挑选一对最优的前景和背景像素作为颜色采样计算初始的  $\alpha$  图; b) 沿着已知前背景区域的边界传播采样点  $\alpha$  值, 以完整捕捉前背景颜色的变化; c) 引入平滑模型, 采用随机游走算法计算全局最优解, 得到精化的结果。

图 4 是 Trimap 和抠图结果的示意图。可以看到生成的 Trimap 较好地给出了分布在前景边缘附近的未知区域, 鲁棒抠图精确地计算出了抠图  $\alpha$  值。

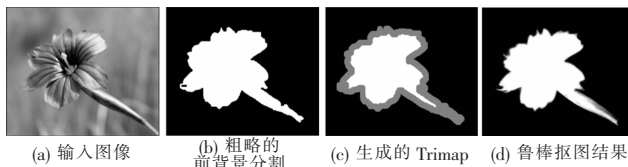


图 4 Trimap 计算和抠图示意图

## 3 实验结果与讨论

本文使用 VC++ 和 MATLAB 在 PC 平台上编程实现了本文算法框架的每个模块。实验的系统配置为 Intel Core™ 2 Quad CPU 处理器, 2.5 GHz 主频以及 3.25 GB 内存。为了证明本文算法框架的实用性和准确性, 本文在标准的抠图数据集和实拍图像上分别进行了实验。对  $400 \times 300$  大小的图片, 本文算法总平均耗时约为 9 s。

### 3.1 标准数据集上的实验

文献[9]提供了用于评价基于给定 Trimap 的抠图算法的标准数据集, 带有真实的  $\alpha$  值。本文算法框架更用于拍摄光圈较大、景深较小的情况, 因此本文选取了其中的一些背景相

对虚化的图像进行实验。图 5 给出了部分实验结果。

在半自动抠图方法中标定线条是一种需要手动操作最少的人工信息提供方式。本文提出的算法是完全只从原始输入图像生成 Trimap 并完成抠图, 因此这里将本文算法的结果与手动描画前背景线条的抠图结果进行了对比。表 1 是实验结果的定量分析。根据文献[9]的评价标准, 本文根据  $\alpha$  值的绝对误差和(SAD)与  $\alpha$  图的均方根误差(MSE)两个方面来衡量结果的准确度, 如表 1 所示。从结果可见, 本文算法获得的  $\alpha$  图更加准确, 而手动标注线条的结果在前景目标的一些区域计算出了错误的  $\alpha$  值, 如图 5(d) 中圈内区域, 导致整体错误率较高。主要原因在于线条标定并不能很好地反映前景和背景完整的颜色分布信息, 而本文的算法生成的 Trimap 可以较好地分割出待计算的未知区域。可见本文的算法框架在不需要人工辅助的同时又能达到很好的精度。

表 1 标准数据集上实验结果的定量比较

方法	图像 GT01		图像 GT22	
	SAD	MSE	SAD	MSE
标注线条方法	24.2	0.201	11.1	0.093
本文方法	1.91	0.035	4.08	0.039

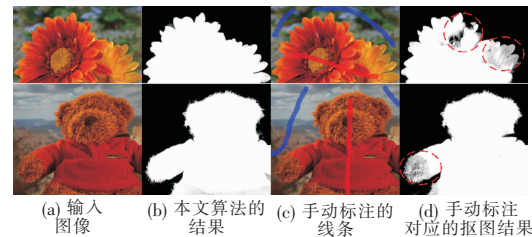


图 5 标准数据集上的本文算法和人工标定线条方式的比较

### 3.2 实拍图片的实验

为了进一步评估本算法的鲁棒性, 本文对一些摄像机采集的实拍图片也进行了验证。图 6 是部分实验结果, 展示了本文算法的中间结果、得到的  $\alpha$  图和使用该抠图结果与新背景合成的新图片。从结果可以看到, 本文算法对玩具毛发等复杂未知区域抠图计算准确, 与其他背景合成的新图像也比较自然逼真。

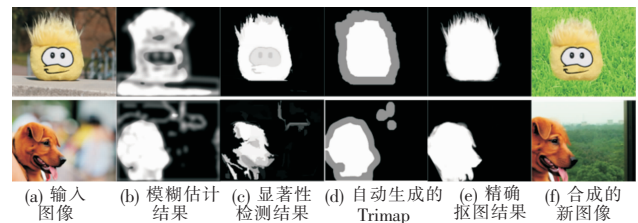


图 6 本文算法在实拍图像上的实验结果

当然本文算法也有一些局限性, 如只适用于较小景深范围的前景聚焦图像或视频, 对前景目标物体的纹理比较弱或是颜色跟背景比较一致的情况可能会失效。后面的工作中会进一步挖掘融合其他可用于辨别深度的图像提示或特征, 引入更加准确的方式提取前景轮廓, 并在算法加速方面展开更深入的研究。

## 4 结束语

本文提出了一种针对自然背景图像的新的自动抠图算法框架。利用散焦模糊度和显著性两个方面的信息进行融合, 区分出前/背景的区域, 自动生成抠图需要的三色图 Trimap, 降低抠图计算的不确定度, 从而得到准确的抠图 (下转第 3955 页)



间,进而提出了基于多帧运动估计的帧率提升系统方案。实验结果表明,与现有常用的提升算法相比,无论是客观指标,还是主观质量,本文算法都具有显著优势,且能保证较好的实时性。

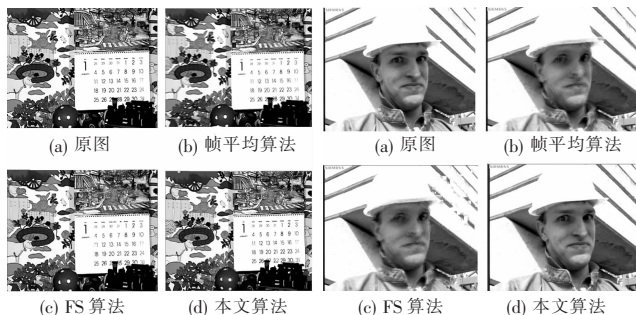


图 10 不同算法内插出 mobile 序列的第 47 帧

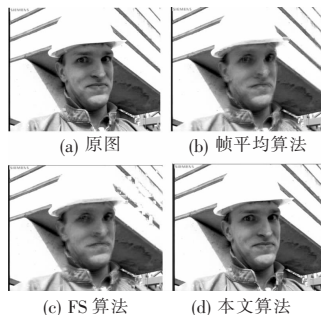


图 11 不同算法内插 foreman 的第 100 帧

### 参考文献:

- [1] BLANCHÉ P, VINCENT A, WANG D, *et al.* Advanced frame rate conversion and performance evaluation [J]. *The SMPTE Motion Imaging Journal*, 2006, 115(4): 153-159.
- [2] TEKALP A M. Digital video processing [M]. 2nd ed. [S. l.]: Prentice Hall, 1995.
- [3] 吴通. 基于 H.264 块匹配运动估计的研究 [D]. 武汉: 武汉理工大学, 2008.
- [4] LI Ren-xiang, ZENG Bing, LIU Ming. A new three-step search algorithm for block motion estimation [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 1994, 4(4): 438-442.
- [5] SHAN Zhu, MA Kai-kuang. A new diamond search algorithm for fast block-matching motion estimation [J]. *IEEE Trans on Image Processing*, 2000, 9(2): 287-290.
- [6] CHOI B T, LEE S H, KO S J. New frame rate up-conversion using bi-directional motion estimation [J]. *IEEE Trans on Consumer Electronics*, 2000, 46(3): 603-609.
- [7] CHEN Tao. Adaptive temporal interpolation using bi-directional motion estimation and compensation [J]. *IEEE Trans on Image Processing*, 2002, 2(9): 313-317.
- [8] TOURAPIS H Y C, TOURAPIS A M. Fast motion estimation within the H.264 codec [J]. *Proc of International Conference on Multimedia and Expo*, 2003, 3(7): 517-520.
- [9] BIERLING M. Displacement estimation by hierarchical block matching [C]//Proc of SPIE Conference on Visual Communication and Image Processing. 1988: 942-951.
- [10] KIM J W, LEE S U. On the hierarchical variable block size motion estimation technique for motion sequence coding [J]. *Optical Engineering*, 1993, 33(8): 2553-2561.
- [11] BURT P J, ADELSON E H. The Laplacian pyramid as a compact image code [J]. *IEEE Trans on Communication*, 1983, 31(4): 532-540.
- [12] WANG De-min, ZHANG Liang, VINCENT A. Motion-compensated frame rate up-conversion-part 1: fast multi-frame motion estimation [J]. *IEEE Trans on Broadcasting*, 2010, 56(2): 133-139.
- [13] KONRAD J, DUBOIS E. Bayesian estimation of motion vector fields [J]. *IEEE Trans on Image Processing*, 1992, 14(9): 910-927.
- [14] CHANG M M, TEKALP A M, SEZAN M I. Simultaneous motion estimation and segmentation [J]. *IEEE Trans on Image Processing*, 1997, 6(9): 1326-1333.
- [15] MCNERNEY P J, KONRAD J. Block-based MAP disparity estimation under alpha-channel constraints [J]. *IEEE Trans on Circuits System Video Technology*, 2007, 17(6): 785-789.
- [16] SHEN Huan-feng, ZHANG Liang-pei, HUANG Bo, *et al.* A MAP approach for joint motion estimation, segmentation, and super resolution [J]. *IEEE Trans on Image Processing*, 2007, 16(2): 479-490.
- [17] WU Jue, CHUNG A C S. A segmentation model using compound Markov random fields based on a boundary model [J]. *IEEE Trans on Image Processing*, 2007, 16(1): 241-252.
- [18] FEGHALI R. Multi-frame simultaneous motion estimation and segmentation [J]. *IEEE Trans on Consumer Electronics*, 2005, 51(1): 245-248.
- [19] GIBSON D, SPAN M. Robust optical flow estimation based on a sparse motion trajectory set [J]. *IEEE Trans on Image Processing*, 2003, 12(4): 431-445.
- [20] INCE S, KONRAD J. Occlusion-aware optical flow estimation [J]. *IEEE Trans on Image Processing*, 2008, 17(8): 1433-1451.
- [21] TEKALP M. Digital video processing [M]. Englewood Cliffs, NJ: Prentice Hall, 1995.
- [22] 宋锦. 帧速率上转换算法的研究与实现 [D]. 天津: 天津大学, 2008.
- [23] 文俊, 刘重庆, 刘晓松. 视频标准帧速率上变换的自适应运动补偿方案 [J]. *数据采集与处理*, 2003, 18(2): 204-208.
- [5] WANG O, FINGER J, YANG Qing-xiong, *et al.* Automatic natural video matting with depth [C]//Proc of the 15th Pacific Conference on Computer Graphics and Applications. Washington DC: IEEE Computer Society, 2007: 469-472.
- [6] BAE S, DURAND F. Defocus magnification [J]. *Computer Graphics Forum*, 2007, 26(3): 571-579.
- [7] CHENG Ming-ming, ZHANG Guo-xin, MITRA N J, *et al.* Global contrast based salient region detection [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2011: 409-416.
- [8] WANG Jue, COHEN M F. Optimized color sampling for robust matting [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2007: 1-8.
- [9] RHEMANN C, ROTHER C, WANG Jue, *et al.* A perceptually motivated online benchmark for image matting [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2009: 1826-1833.

(上接第 3947 页)结果。本文算法的优点在于完全自动进行,无须人工辅助或附加信息。实验结果证明了本文算法的有效性与准确性。未来将进一步融合更多的深度提示信息,扩大算法的应用范围。

### 参考文献:

- [1] WANG Jue, COHEN M F. Image and video matting: a survey [J]. *Foundations and Trends in Computer Graphics and Vision*, 2007, 3(2): 97-175.
- [2] McGUIRE M, MATUSIK W, PFISTER H, *et al.* Defocus video matting [C]//Proc of ACM SIGGRAPH. New York: ACM Press, 2005: 567-576.
- [3] SUN Jian, LI Yin, KANG S B, *et al.* Flash matting [C]//Proc of ACM SIGGRAPH. New York: ACM Press, 2006: 772-778.
- [4] JOSHI N, MATUSIK W, AVIDAN S. Natural video matting using camera arrays [C]//Proc of ACM SIGGRAPH. NY, USA: ACM Press, 2006: 779-786.