

数字抠像的最新研究进展

张展鹏^{1,2} 朱青松¹ 谢耀钦¹

摘要 数字抠像是图像处理、视频编辑和电影制作中的关键技术。通过数字抠像，从图像或视频的背景中精确地分离出前景，是计算机视觉领域的重要问题。本文首先介绍了目前数字抠像的交互方式，然后把抠像技术分为基于颜色采样、基于像素相似性、基于能量函数以及基于机器学习的四类技术，介绍和分析了其中的典型算法和最新研究成果，并对这些算法的测试结果进行了定量和定性比较，最后总结了数字抠像技术目前的研究状况和未来的发展方向。

关键词 数字抠像，自然图像抠像，图像分割，图像处理

引用格式 张展鹏, 朱青松, 谢耀钦. 数字抠像的最新研究进展. 自动化学报, 2012, 38(10): 1571–1584

DOI 10.3724/SP.J.1004.2012.01571

The Latest Research Progress on Digital Matting

ZHANG Zhan-Peng^{1,2} ZHU Qing-Song¹ XIE Yao-Qin¹

Abstract Digital matting is the key technology in image processing, video editing and film-making applications. It refers to the problem of extracting the foreground objects in the images or videos accurately, which is an important issue in the field of computer vision. In this paper, the interaction modes for digital matting are introduced firstly. After that, matting techniques are divided into four categories: color sampling based, pixel affinity based, energy function based and machine learning based. Classic algorithms and the latest research progress are presented and analyzed, followed by quantitative and qualitative evaluations of these techniques. Finally, the research progress is summarized and future research directions are suggested.

Key words Digital matting, natural image matting, image segmentation, image processing

Citation Zhang Zhan-Peng, Zhu Qing-Song, Xie Yao-Qin. The latest research progress on digital matting. *Acta Automatica Sinica*, 2012, 38(10): 1571–1584

数字抠像 (Digital matting) 是指从图像或视频的背景中精确地分离出前景，是图像处理、视频编辑和电影制作中的关键技术，已得到广泛的研究和应用。抠像过程中，设图像 I 包含 N 个像素，即 $I = \{I_1, I_2, I_3, \dots, I_N\}$ ，其中每个像素 I_z 的颜色 C_z 可表示成其前景色 F_z 与背景色 B_z 的线性组合，即

$$C_z = \alpha_z F_z + (1 - \alpha_z) B_z \quad (1)$$

下文将式 (1) 称为抠像方程。其中 $\alpha_z \in [0, 1]$ ，表示像素 I_z 在图像中的前景不透明度。若 $\alpha_z = 1$ ，则该像素属于前景；若 $\alpha_z = 0$ ，则属于背景。若 $0 < \alpha_z < 1$ ，则该像素为前景与背景的混合，可称为“混合像素”。混合像素一般出现在半透明的物体或毛绒状

收稿日期 2011-10-08 录用日期 2012-05-10

Manuscript received October 8, 2011; accepted May 10, 2012
国家重点基础研究发展计划 (973 计划) (2010CB732606), 国家自然科学基金 (81171402, 30928030) 资助

Supported by National Basic Research Program of China (973 Program) (2010CB732606) and National Natural Science Foundation of China (81171402, 30928030)

本文责任编辑 戴琼海

Recommended by Associate Editor DAI Qiong-Hai

1. 中国科学院深圳先进技术研究院 深圳 518055 2. 中山大学 广州 510275

1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055 2. Sun Yat-Sen University, Guangzhou 510275

的前景边缘，图像离散化过程中造成的非连续性、运动或光照带来的模糊是形成混合像素的主要原因。

对于彩色图像 I ，需要为每个像素 I_z 估计前景色、背景色以及前景不透明度。已知该像素的颜色 C_z ，若采用 RGB 颜色空间表示，则 C_z, F_z, B_z 均为三维向量，对于式 (1) 可得出 3 个方程，3 个已知变量和 7 个未知变量。显然这是不定方程，即病态 (Ill-posed) 问题^[1]。因此，通常需要使用先验假设或用户提供的额外信息，为抠像问题增加限制条件，然后进行求解。灰度图像与此类似。

早期研究中，通过把前景物体置于已知颜色的背景前采集图像，从而减少方程中的未知数，这种方法被称为“蓝屏抠像”。Smith 等在 1996 年就提出一种三角抠像法^[2]。该方法把同一前景物体置于多个不同的背景中，得到多幅图像，再抠选前景。因为这些背景的颜色已知，增加了抠像问题中的已知信息，使得抠像方程有确定的解。由于三角抠像把抠像问题转化为求解超定方程，且抠像效果良好，在目前的研究中，常被用于生成标准的抠像结果 (Ground truth)^[3-4] (如图 1(d))，作为算法测试和评价的依据。然而，由于蓝屏抠像需要已知且固定的背景，因此应用范围不广。

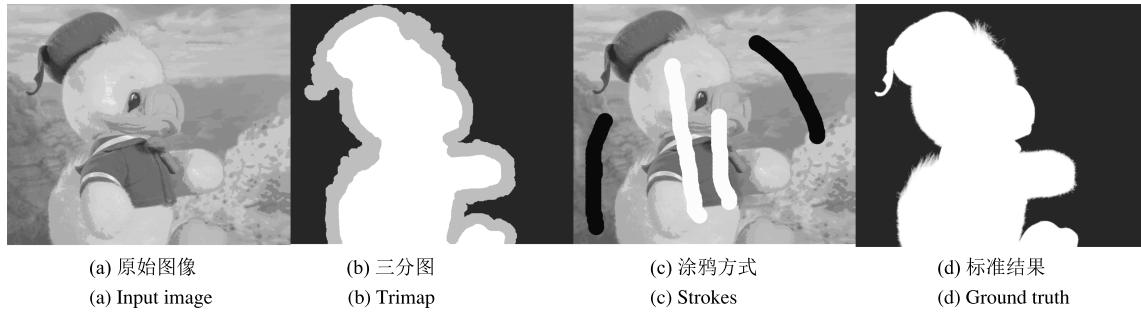


图 1 数字抠像的交互方式

Fig. 1 Interaction modes in digital matting

近年的研究集中在自然图像抠像, 即不对背景进行限制, 面向任意的自然图像。此外, 数字抠像的扩展应用, 如环境抠像、阴影抠像以及视频抠像, 也受到关注^[5]。环境抠像除了精确地提取前景以外, 还需要得到物体对光照的反射和折射特性, 使得前景物体在新合成的图像中也能体现这些性质^[6]。这种技术对于透明物体的抠像^[7]特别重要。阴影抠像是把图像中的阴影部分提取出来, 去除原图像的阴影^[8]或把阴影合成到新的背景^[9]。视频抠像可看作数字抠像在连续图像序列的扩展, 有效的物体跟踪^[10]和运动分割^[11]能简化用户的操作。

由于自然图像抠像是近年的研究热点、具有实用价值且应用广泛, 因此下文的论述中将以自然图像抠像为重点展开, 介绍数字抠像最新的研究进展。同时为了内容的全面性及完整性, 也会详细分析一些经典算法。

1 数字抠像的交互方式

由于数字抠像是一个病态问题, 需要获得额外的信息进行求解, 因此, 在目前的算法中, 经常通过用户的交互, 获得更多输入信息, 构建约束条件。主要的交互方式有三分图和涂鸦方式。

1.1 三分图

三分图 (Trimap) 是指一幅大小与原图像相等的图像, 图像被用户划分为前景区、背景区以及未知区域。在这种情况下, 前景/背景区为已知区域, 抠像算法需要估算未知区域像素的前景色、背景色以及前景不透明度。图 1(b) 为一幅三分图, 其中背景标注为黑色, 前景为白色, 灰色为未知区域。三分图是对原图像的粗略划分, 是自然图像抠像的研究中, 最开始采用的输入方式。文献 [3, 12] 等均采用三分图作为算法的输入。然而, 创建一幅三分图往往需要较多的用户操作, 对于一些形状复杂的图像(如蜘蛛网), 创建的工作非常困难。Soft scissors^[13] 中实现了一种智能的描边笔刷, 能够根据实时的描边情况改变笔刷大小, 降低边界定位的难度, 然而在前景边

界较长或形状复杂的情况下, 这种工作仍然繁琐。

1.2 涂鸦方式

由于生成三分图的工作比较困难和繁琐, 为了提高数字抠像的实用性, 越来越多抠像技术采用涂鸦方式^[14–16], 以提供良好的用户体验。如图 1(c), 用户只需通过涂鸦式的操作, 使用笔刷, 在前景和背景的其中一小部分上做标记。由于涂鸦的结果可看作三分图的子集, 一般支持涂鸦式交互的技术也支持三分图。涂鸦方式下, 用户的输入更为简单。然而, 由于只对某部分进行了标记, 没有得到较为完整的前景和背景样本, 难以对大面积的未知区域进行估算。因此, 一些算法先进行图像分割, 自动生成大致的三分图^[17], 或者从标记的已知区域开始, 迭代式地对附近的像素进行估算, 逐渐增大已知区域直到抠像完成^[18]。也有算法把用户的标记转化为求解不定方程的限制条件^[12, 16]。还有采用迭代式的涂鸦, 即根据用户的每次涂鸦操作, 对结果进行改进。文献 [14] 中实现了一种基于涂鸦输入的交互式抠像方法, 每次用户添加了新的涂鸦标记后, 算法只需更新图中的部分像素而不需重新全部计算, 从而提高响应速度。然而, 当用户涂鸦出错想擦去时, 则没有应对的方案, 或者需要重新对全部像素进行计算。此外, 由于涂鸦操作获得的初始样本较少, 实际中算法不一定能得到有代表性的样本, 而且抠像效果容易受到图像噪声的影响。

除了三分图和涂鸦方式, 也有其他的交互方式。比如在 GrabCut^[19] 中, 用户使用一个矩形框选中图像中的前景部分。在很多情况下, 这种方式更为直观。但是, 由于选中的区域中也包含了部分背景, 算法不能获得准确的前景样本, 难以确定前景边界, 需要用户进行一些边界标注。还有方法首先通过无监督的方式, 根据像素间的相似性, 自动地对图像进行区域分割, 然后指导用户在需要提取的区域上点击或做上记号, 接着通过这些标记信息对各个区域进行合并或提取, 继续完成抠像^[20–21]。

上述方法中, 虽然用户操作的复杂性不同, 但共

同的特点是, 用户需要对图像中的某些区域进行标记, 指明前景区或背景区, 因此属于半自动的方式. 而一些研究中致力于实现全自动抠像. 例如, 闪光(Flash) 抠像^[22] 通过对同一场景进行两次拍摄, 分别开启和关闭闪光灯, 得到两张图像作为算法的输入. 计算过程中取两张图像的差值, 近似地提取出被闪光的前景, 然后进行自动抠像. 另外, 立体图像(Stereo image)^[23]、同一场景下不同焦点平面的多幅图像^[24] 也被作为算法输入, 实现自动抠像.

2 数字抠像的主要技术

2.1 基于颜色采样的技术

普遍图像中, 相近的像素在统计特征上往往具有相关性, 可以对相近的像素进行颜色采样, 根据样本颜色的特点对未知区域像素的抠像参数(F, B, α)进行估算. Berman 等^[25] 对周边确定区域像素的前景/背景色进行加权, 作为未知区域像素的估算结果. Ruzon 等^[26] 则最先在数字抠像中引入概率统计, 其基本思想是: 对于未知区域的像素, 先取附近已知的前景和背景像素作为样本, 进行聚类和统计, 每个聚类使用高斯模型进行描述. 然后根据样本颜色的概率模型和未知区域像素的颜色, 估算未知像素与前景/背景中各个聚类的相似度, 推导该像素的前景不透明度. 2001 年, Chuang 等^[3] 在此基础上提出了贝叶斯抠像(Bayesian matting). 这种方法使用贝叶斯公式, 把未知像素的估算问题转化为一个最大后验概率问题. 算法已知该像素的颜色 C , 目标是通过估计前景色 F 、背景色 B 以及 α 值, 最大化后验概率, 如式(2)所示:

$$\begin{aligned} & \arg \max_{F, B, \alpha} P(F, B, \alpha | C) = \\ & \arg \max_{F, B, \alpha} \frac{P(C|F, B, \alpha)P(F)P(B)P(\alpha)}{P(C)} \end{aligned} \quad (2)$$

其中概率 $P(\cdot)$ 均使用高斯分布模型进行描述. 其中 $P(\alpha)$ 看作常数. $P(C|F, B, \alpha)$ 使用 $\alpha F - (1 - \alpha)B$ 作为高斯分布的均值. 而对于 $P(F), P(B)$, 在未知像素附近采集前景/背景样本, 分别进行聚类. 求出每个聚类中的均值和协方差矩阵作为高斯分布的参数, 求解 F, B 对应的概率. 由于抠像过程中, 需要对每个未知像素的样本进行聚类和统计, 因此计算量较大. 而且, 算法把未知区域看成前景区与背景区夹着的带状区域, 抠像时, 从带状区域两侧开始计算, 逐步向中间部分推进, 计算中会利用之前已求解的像素作为样本, 所以求解过程中的误差会被累积, 这种现象在带状区域较宽的情况下更为明显. 不过, 这种算法已取得当时最好的效果, 所以得到广泛的参考和改进, 是数字抠像的经典算法. 例如,

文献[18] 在求解 $P(F), P(B)$ 时, 没有使用统计模型对样本进行匹配或聚类, 而仅仅基于空间距离和算法对样本的“置信度”, 进行前景/背景色的估算, 从而达到减少计算量的目的; 文献[27] 则分别对前景和背景区域使用全局的高斯混合模型(Gaussian mixture model, GMM)进行统计建模, 减少采用局部统计带来的计算量. 然而, 这些方法仍未解决基于颜色采样的技术中普遍存在的问题, 即在前景和背景颜色接近、具有相似的统计特征或者色彩模糊的情况下, 效果往往不佳. 其主要原因是采集的样本只是图像中相近的像素, 这些像素不一定能够有效地表示出未知像素的特征.

为了提高颜色采样的效果, 采样技术成为了近年来数字抠像的研究热点^[28–31]. Wang 等在 2007 年提出 Robust matting^[28] 方案, 方案中根据“信任系数”的大小判断一对前景/背景像素能否作为样本. 信任系数的计算主要考虑样本与未知像素间的颜色空间距离. 而对于候选择本集的构建, 与之前的方法不同, 不仅选取和未知像素相近的点, 同时沿着附近已知区域的边缘进行扩张. 通常, 抠像中需要处理很多条状的未知区域(例如头发), 这种方法可以取得距离更远和种类更丰富的样本, 而且样本的颜色也与未知像素较为接近. 但有些情况下, 三分图中的未知区域较宽, 而且已知区域的边缘处, 其形状、颜色与未知区域也不吻合, 这种方法得到的样本并不具有代表性. Rhemann 等^[29] 指出, 测地线距离能够反映图像中的形状信息, 同一形状上的像素关联度更大, 因此可以用测地线距离进行辅助, 建立候选择本集. 如图 2(a) 中, F 和 B 区域分别是三分图中指定的前景区和背景区, 灰色和白色的采样点分别表示 Wang 和 Rhemann 的方法得到的结果. 可以看出, 灰色的点与未知像素(即图 2(a) 中的点 i) 分布在同一物体上, 颜色特征会更接近. 然而, 上述两种方法都是从邻近像素出发, 进行扩张的, 要利用位置距离更远的像素, 则需要遍历较大的空间, 从而带来很大的计算量. 2010 年, Gastal 等在其 Shared matting^[30] 方案中, 通过相邻像素间共享候选择本来减少计算开销. 而且, 在采集候选择本的时候, 从每个像素出发, 沿四条射线向外寻找, 从而扩大采集范围(如图 2(b)). He 等^[31] 在 2011 年的最新成果中, 进一步设计了一个全局的采样方法. 三分图中的所有前景 F 和背景 B 的像素组成一个大小为 $N_F \times N_B$ 的矩阵, 作为候选择本的“ FB 搜索空间”. 然后, 把候选择本的采集看作图像像素与 FB 搜索空间中样本对的匹配问题(如图 2(c)), 匹配的标准基于颜色空间距离以及像素位置距离. 为了避免对 FB 搜索空间的穷举, 把样本的搜索算法分为两步: 扩张搜索与随机搜索. 算法的运行过程中, 交替地执行这两个步骤, 添加符合的样本: 在扩张搜索

中, 根据邻近像素相似, 考虑当前最佳样本的邻近节点; 在随机搜索中, 则考虑一定范围内随机选定的样本对。由于采用了全局的采样方法, 可以有效避免合适样本的丢失, 同时, 使用扩张搜索和随机搜索相结合的方式, 避免了穷举搜索带来的巨大计算量。

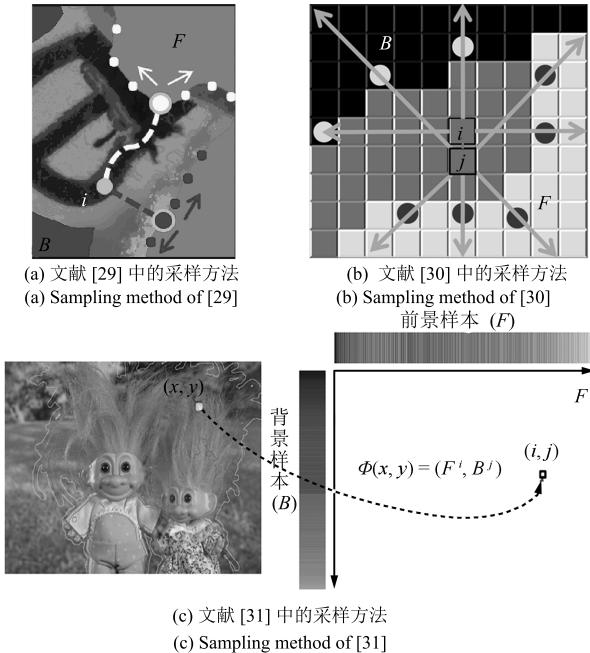


图 2 抠像中的多种采样方法^[29–31]

Fig. 2 Various sampling methods in image matting^[29–31]

除了前景/背景的采样技术, 颜色模型方面也有相应的研究。Lin 等^[32] 使用感知颜色空间替换 RGB 颜色空间进行建模, 把透明度的计算细分为亮度和色度透明度的计算, 提高计算的精确度。Cho 等^[33] 的方法则关注未知像素和附近样本颜色间的关系。基于二次贝塞尔 (Bezier) 曲线, 引入了一个自适应的颜色曲线模型, 对局部区域中各像素的颜色进行描述。通过计算未知像素的颜色在曲线中的位置, 得出抠像参数。

总体来说, 基于颜色采样的技术, 在不同的图像中, 抠像结果的质量差异很大^[1]。在三分图被仔细定义、前景/背景对比明显的情况下会取得良好的效果。在纹理或色彩复杂的情况下, 样本选取不当会对抠像结果产生不良的影响。为了提高采样统计的有效性, 近年来研究学者做了各种努力。然而, 由于目前候选样本的采用标准只考虑位置和颜色距离, 仍然不能避免一些原本相关性不大的样本获得较高的信任系数, 特别是在一些颜色模糊的区域。

2.2 基于像素相似性的技术

基于像素相似性的技术假设在某种距离 (例如位置距离、颜色空间距离) 下, 邻近的像素具有相似的属性或符合一定的规律^[12,14,34–35], 例如, 假设相

邻像素间的前景/背景色平滑过渡^[12]、具有相同的模糊连接度 (Fuzzy connectedness)^[14]。与颜色采样的技术相比, 所用的邻近像素较少, 可以充分利用邻近像素的相关性, 减少由于颜色采样中样本选取不当导致的错误。而且基于邻近像素的相似性, 可以促使求解出的前景不透明度在像素间平滑过渡, 减少颜色采样技术中可能产生的不连续问题, 有利于提高视觉效果。

这类技术中最典型的是泊松抠像^[12], 由 Sun 等在 2004 年国际图形学会议 (SIGGRAPH) 论文中提出。该算法把抠像问题转化成求解关于 α 梯度的泊松方程。首先对抠像方程的左右两边求偏导, 得:

$$\nabla C = (F - B)\nabla\alpha + \alpha\nabla F + (1 - \alpha)\nabla B \quad (3)$$

其中, $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ 。算法中假设图像中的前景色和背景色平滑过渡, 因此 $\alpha\nabla F + (1 - \alpha)\nabla B$ 相对较小, 可以忽略。 α 值的梯度 $\nabla\alpha$ 约等于 $\frac{\nabla C}{F - B}$ 。然后利用三分图中的区域划分, 已知区域边缘像素的 α 值定为 0 或 1, 从而确定了狄里克雷边界条件, 可以求解对应的泊松方程 (如式 (4)), 得到 α 值。

$$\Delta\alpha = \text{div}\left(\frac{\nabla C}{F - B}\right) \quad (4)$$

其中, $\Delta = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ 为拉普拉斯算子, div 表示散度。接着, 使用迭代的方法进行求解, 即不断地优化 F, B 的取值以使得 α 的变化足够小。由于 F, B 的初始值从最近的点选取, 而且迭代过程中没有有效的信息对 F, B 进行优化, 只是基于图像“平滑过渡”的假设, 因此, 在一些颜色复杂的图像中, 这不能取得好的结果。为此, Sun 等也提出了局部泊松抠像, 即通过用户对局部区域的操作, 为全局抠像中忽略的部分进行赋值, 加入到计算中, 在一些背景较复杂的图片中取得了比贝叶斯抠像更好的结果。然而, 局部泊松抠像需要较多的用户交互, 整个抠像过程时间较长。Du 等^[36] 对此进行了改进, 提出使用一系列的滤波操作提取图像的细微特征, 把这些特征也作为抠像算法的输入。同时, 把原来基于散度的泊松方程 (式 (4)) 改写成基于特征值的方程, 使得方程求解中更好地体现图像细节。通过这些方法, 用户只需提供初始的三分图而不需进行进一步的调整操作, 提高了抠像效率。

由于测地线距离可以反映图像中物体的形状信息, 基于测地线的抠像方法近几年也受到关注^[34,37–38]。Bai 等^[34] 提出基于测地线的抠像框架。其基本思想是使用像素间的测地线距离, 结合前景和背景的颜色概率分布, 求出未知像素与前景的相似度, 从而得出 α 值。像素间的测地线距离定义为

$$d(x, z) = \min_{C_{x,z}} \int_0^1 |W \cdot \hat{C}_{x,z}(p)| dp \quad (5)$$

其中, $C_{x,z}$ 是连接像素 x 与 z 之间的路径. 作为权值的 W 表征像素与前景相似度的梯度, 定义为

$$W = \nabla P_F(x) = \nabla \frac{P(x|F)}{P(x|F) + P(x|B)} \quad (6)$$

其中, $P(x|F)$, $P(x|B)$ 分别代表前景和背景颜色的概率分布函数. 而像素 x 到前景和背景区域的测地线距离 $D_F(x)$, $D_B(x)$ 则定义为 x 到前景/背景区的最小测地线距离. 然后, 求解式(7)得到 α 值.

$$\alpha(x) = \frac{w_F(x)}{w_F(x) + w_B(x)} \quad (7)$$

其中, $w_F(x) = D_F(x)^{-r} \cdot P_F(x)$, r 是控制边界平滑程度的参数. $w_B(x)$ 与此类似.

这种方法中, 测地线距离是基于权重距离的函数, 可以在线性时间完成计算. 这个优点在算法扩展到视频抠像的时候更加明显. 然而, 由于权重计算时简单地基于前景和背景的颜色概率模型, 当前景和背景色域重叠时, 权重计算会出现错误. 而且, 这种基于测地线的方法, 对用户指定的样本比较敏感, 容易受噪声影响, 出现边界估计结果不稳定的情况. 此外, 方法中缺少对边界的求解, 当图像实际边界较为明显时可得出正确的结果, 当边界模糊时则会出现问题^[37].

Levin 等^[16] 的方法中, 则对抠像问题求出了闭合解 (Closed-form solution), 成为近来很多研究的基础^[20, 39–43]. 以大小为 $N \times N$ 的灰度图像 I 为例, 通过假定图像中的一小块区域内, 前景色 F 和背景色 B 保持恒定, 然后对抠像方程进行改写, 并定义了一个二次目标函数, 经过进一步的代数运算, 目标函数化简成只与 α 相关的函数, 如式(8)所示:

$$J(\alpha) = \alpha^T L \alpha \quad (8)$$

其中, L 是 $N \times N$ 的矩阵, 矩阵中元素 (i, j) 定义如下:

$$\sum_{k|(i,j) \in w_k} \left(\delta_{ij} - \frac{1}{|w_k|} \left(1 + \frac{1}{\frac{\epsilon}{|w_k|} + \sigma_k^2} \times (I_i - \mu_k)(I_j - \mu_k) \right) \right) \quad (9)$$

其中, $|w_k|$ 为包围像素 k 的窗口区域的大小, δ_{ij} 为克罗内克函数, μ_k , σ_k^2 分别为 w_k 内灰度值的均值和方差. 然后通过最小化 $J(\alpha)$, 可得出 α 的值. 对于彩色图像, 则假定像素颜色符合 Color line 模型, 即在一小块区域内, 前景色 F 或背景色 B 可以表示为

某两种颜色的线性组合, 区域内像素的 F 和 B 分布在 RGB 空间的同一直线上. 基于这个假定, 可以推导出类似上述的结果. Levin 等的这种方法通过添加一个较弱的假设, 结合巧妙的代数变换, 最终把抠像方程这个不定问题转化为确定问题, 并求出了闭合解.

Closed-form matting 中推导出的矩阵 L 被称为“拉普拉斯抠像矩阵”, 是 Levin 等重要的理论成果, 往后被广泛地引用. 取 L 的若干个最小特征值对应的特征向量进行聚类, 从而可把图像分成多个区域, 每个区域内的像素具有较大的相似性. 文献[20]在此基础上, 利用谱聚类的思想, 实现了谱抠像 (Spectral matting). 首先使用拉普拉斯抠像矩阵, 无监督地对图像进行区域划分, 然后初步组合出前景和背景. 用户也可以对图像中的区域进行标注, 指明前景和背景, 根据这些限制信息, 算法继续优化得出最终结果. 可是当图像中纹理比较复杂时, 图像将被分割成很多部分, 如何进行组合是一个问题. He 等^[39] 则在计算效率方面对拉普拉斯抠像矩阵进行了探讨. 由于拉普拉斯抠像矩阵行列数都等于像素数目, 随着图像增大, 计算量会剧增. 以前的方案中, 一般认为使得拉普拉斯矩阵越稀疏, 就越能减少计算量, 即式(9)中的窗口大小 $|w_k|$ 倾向于取很小的值. 而 He 等则指出这种直觉不一定正确, 并实现了一种算法, 在使用较大窗口的情况下, 通过减少求解中的迭代次数, 反而取得了更高的计算效率. 另外, Chen 等^[44] 在抠像方程中加入两个中间变量, 对方程进行改写, 添加像素间的平滑假设, 抠像问题转化成一个线性规划问题, 同样得到了闭合解.

点扩散函数 (Point spread function, PSF) 也被用于数字抠像. 点扩散函数描述光学系统接收一个点光源后, 在成像区出现的区域光场. 图像中一些“混合像素”(即 $0 < \alpha < 1$ 对应的像素) 由这种现象产生, 因此, PSF 描述了局部区域内各像素 α 值的相互关系, 可以把 PSF 作为先验知识进行抠像. Rhemann 等^[17] 在 2008 年最先提出这个方法. 整个图像的 α 平面, 有以下定义:

$$\alpha = K \otimes \alpha^s \quad (10)$$

其中, K 是 PSF 对应的矩阵形式, \otimes 是卷积运算, α^s 是受 PSF 影响前的值. 在 2010 年的国际计算机视觉与模式识别 (Conference on computer vision and pattern recognition, CVPR) 会议论文中, Rhemann 等继续对 PSF 用于抠像的方法进行了优化^[45]. 首先用现有的抠像算法求出整个图像的 α 平面, 然后对 α 平面进行上采样 (Upsampling), 得到更高的分辨率, 接着估计出二值的 α^b , 即不透明度等于 0 或 1. 把图像根据模糊程度划分成若干区域, 估算每个区域对应的 PSF. 使用得到的 PSF 与 α^b

构建出 α^{prior} 作为先验知识, 最后代入现有的抠像算法中进行计算.

从泊松抠像开始, 学术界就陆续提出了不少基于像素相似性的抠像方法. 这些方法基于不同的相似性假设, 得到先验知识, 把抠像问题转换为数学上有确定解的问题. 由于不涉及到采样统计, 只采用邻近的像素, 充分利用邻近像素的相关性, 这类算法在一些色彩稍复杂的图像中, 可以取得比颜色采样算法更好的结果. 然而不同的假设直接影响到算法对不同图像的适应性. 例如, 泊松抠像假设前景/背景色变化不大, 因而在颜色变化频繁的图像中会出现较多错误; 测地线距离的方法受用户输入的影响较大; 应用 PSF 的方法受限于混合像素产生的原因, 且增加了额外的计算量.

2.3 基于能量函数的技术

由于基于颜色采样和像素相似性的技术各有优缺点和适用范围, 一些学者把两种思想融合到一个能量函数, 同时描述当前抠像方案下, 区域内统计上的相容性和像素间的渐变性, 以增加算法的健壮性. Boykov 等^[46] 把图像边界分割问题定义为一个图割 (Graph cut) 问题. 图像被看作一个图 (Graph), 像素即其中的节点. 图像分割的目标转换成寻找一个“最小割”, 对图像进行前景/背景划分. 具体地, 需要最小化一个预设的能量函数, 即

$$E(A) = \lambda R(A) + B(A) \quad (11)$$

其中, A 表示当前的划分方案, 指定每个像素属于前景还是背景. λ 为权重参数. $R(A)$ 表征区域性质, 例如与前景颜色相近的点应该偏向于划分为前景, 这样可使 $R(A)$ 的值偏小. $B(A)$ 表征边界性质, 促使颜色差异较大的地方划分为边界. 这种图割思想后来被引入到数字抠像. 学者使用各种方法对 $R(A)$ 、 $B(A)$ 进行具体的定义. $R(A)$ 一般被称为数据项. $B(A)$ 被称为边界项、平滑项或邻接项, 名称不同而作用大体一致. Lazy snapping^[47] 以及 GrabCut^[19] 都是这类算法的典型代表.

GrabCut 在 Boykov 算法的基础上改进而成. 其方案分为两步: 首先进行图像的二值分割, 即划分出前景区和背景区, 然后对边界像素的 α 值进行平滑处理. 第一步中, 对式 (11) 的各项进行了重新的定义. 数据项 R 表征 GMM 模型下邻近像素的相容性, 边界项 B 表征区域间像素的颜色空间距离. 邻近像素间越符合 GMM 模型 (概率值越大) 则数据项越小, 不同区域间像素的颜色距离越大则边界项越小. 算法通过一个迭代的过程, 调整划分方案, 最小化能量函数, 则可得到二值分割的结果.

接着, 进行第二步, 对边界像素进行 α 值的平滑处理. 以第一步中划分出的前景/背景分割线为中

心, 取一定的宽度, 划分出一个带状区域, 作为未知区域 U , 从而得到新的三分图. 算法假设 α 值在边界处渐变, 使用一个软阶跃函数描述 α 的变化. 新的能量函数定义如式 (12) 所示:

$$E = \sum_{n \in U} D_n(\alpha_n) + \sum_{t=1}^T V(\Delta_t, \sigma_t, \Delta_{t+1}, \sigma_{t+1}) \quad (12)$$

算法把带状未知区域 U 分成 T 小段, 对每段进行计算. Δ 和 σ 是描述软阶跃函数的参数. 其中边界项 V , 表征从第 t 段到第 $t+1$ 段参数值的变化大小. 相邻分段间参数变化越小, 则 α 值变化越小. 至于数据项 D , 算法对区域 U 附近的前景/背景样本进行采样统计, 建立高斯分布模型. 与二值划分时类似, 在所建立的概率模型下, D 的值与 α_n 取值下的概率负相关. 因此, 最小化能量一方面会使得 α 的取值与邻近的区域相容, 一方面又要求 α 能平滑地变化. 因为 U 被截成有限小段, 可使用动态规划进行求解. 然而, 由于使用软阶跃函数描述 α 的变化, 算法中对边界的平滑处理实际上是一个羽化的过程, 这种方法并不适合于一些边缘复杂的情况.

Wang 等^[48] 则构建了一个马尔科夫随机场 (Markov random field) 来解决抠像问题. 把待估算的像素看做场内的节点, 分别与 4 邻域的节点相连接. 同时对 α 值进行离散化处理, 分为 25 个可能的取值, 从而可以用 α 值表示节点所处的状态. 同样地, 定义场的能量中包含数据项和邻接项, 然后通过信任传播 (Belief propagation, BP) 算法最小化能量函数, 得到 α . 然而, BP 算法收敛速度较慢, 影响了抠像效率; Guan 等^[15] 则在能量函数中引入了动态的权重, 调节数据项和平滑项的影响力. 其算法基于这样的假设: 在前景的内部, 物体一般较为平滑, 此时增大平滑项的权重, 使其在能量函数中发挥更大的影响. 在边缘处, 颜色对比较为明显, 减少平滑项的权重, 更多地利用数据项的优势. 此外, 和泊松抠像的方法类似, 采用全局抠像和局部抠像相结合的方法. 用户可以对某些区域调节其平滑项的权重, 或者为某些区域的前景色选择一些对比明显的背景样本, 增加数据项的有效性.

在 2007 年的 CVPR 会议论文中, Wang 等又提出了新的方案, 即 Robust matting^[28]. 与之前基于能量函数的方法相比, 其主要思想是对数据项和平滑项的权重有了更精确的调节. 如第 2.1 节所述, 首先根据颜色空间距离, 得出样本的信任系数 \hat{f} , 再直接根据样本粗略估计出对应的不透明度 $\hat{\alpha}$. 然后最小化能量函数, 能量函数定义如式 (13) 所示:

$$E = \sum_{z \in \psi} \left[\hat{f}_z (\alpha_z - \hat{\alpha}_z)^2 + \right.$$

$$(1 - \hat{f}_z)(\alpha_z - \delta(\hat{\alpha}_z > 0.5))^2] + \lambda \cdot J(\alpha) \quad (13)$$

其中, δ 表示布尔运算, 返回 0 或 1. 对左边项稍作直观分析, 当样本信任系数高, 则认为估计算的 $\hat{\alpha}_z$ 可信, 因此使 α_z 与 $\hat{\alpha}_z$ 的差值变小. 而当信任系数较小时, $\hat{\alpha}_z$ 的具体数值相对不可靠, α_z 只需根据一个阀值 (0.5) 大致追随估算值. 右边项 $J(\alpha)$ 的定义与式 (8) 相同, 可见 Robust matting 中融入了 Closed-form matting 的方法. 其原因是 Closed-form matting 中充分利用了邻近像素的相关性, 正好作为平滑项与数据项互补.

此外, 近两年的研究中, 开始偏向于在数据项或平滑项赋予更多的假设或先验知识, 使得抠像更准确. 例如 Park 等^[49] 把 Graph cut 和 Closed-form matting 结合起来, 分别利用两者在“硬”分割和“软”分割上的优势实现抠像. Price 等^[37] 则在能量函数中融入了测地线距离的度量.

2.4 基于机器学习的技术

从数字抠像的已知条件和求解目标来看, 可以把抠像的过程看作建立 α 与图像颜色之间的模型. 在颜色采样的技术中, 通过颜色的相关度来寻找这些相关点, 然后假设像素间的前景/背景色符合某种统计模型 (如 GMM), 求解的目标是使未知像素的抠像参数拟合统计模型; 基于像素相似性的技术与此类似, 通过各种方法定义相邻的点 (如位置相邻、测地线距离相近), 然后假设像素间符合一定的关系 (如平滑过渡、点扩散函数关系), 基于假设条件对抠像方程或目标函数进行代数求解得出结果. 假设条件越弱, 则算法健壮性越强, 精确度越高. 基于机器学习的技术与之不同, 这类算法把抠像的过程看做一个监督或半监督学习的问题, 通过一个学习的过程, 建立 α 与图像颜色之间的模型, 而不依赖于较强的模型假设.

Zheng 等^[50] 在 2009 年的国际计算机视觉 (International conference on computer vision, ICCV) 会议论文中运用了这种基于学习的方法. 算法的目标是通过训练, 建立 α 与图像颜色之间的模型, 从而预测未知像素的 α 值. 模型可以是线性或非线性的. 这里简单介绍其模型的建立方法. 为便于表述, 先介绍线性模型的情况, 再进行扩展. 设局部区域中, α 与图像颜色间的模型可表示为

$$\alpha = x^T \beta + \beta_0 \quad (14)$$

其中, x 是局部区域中各像素的颜色空间向量的集合, β 、 β_0 分别是模型的参数. 其中 β_0 为标量, β 为矢量, 维数等于图像颜色的通道数. 通过岭回归 (Ridge regression)^[51] 技术, 可以把模型参数 β , β_0

的估算转化为一个二次最优化问题, 从而估算出 β , β_0 , 即描述出模型. 对于非线性模型, 使用核方法 (Kernel trick) 进行转换, 把线性情况下, β , β_0 的求解结果中, 包含的向量 x'_i 与 x'_j 的内积, 使用核函数 $k(x_i, x_j)$ 替换, 其中, $x' = [x^T \ 1]^T$. 具体地, 可以使用高斯核, 即 $k(x'_i, x'_j) = \exp(\frac{1}{\vartheta} \|x'_i - x'_j\|^2)$. ϑ 使用图像灰度值的方差. 与之前的方法相比, 这种方法有两个优势: 一是实现简单, 在整个过程中只需一些矩阵操作; 二是针对线性和非线性的模型有相应的处理方法, 建立的模型可以更一般化.

除了 Zheng 等的这个方法, 主成分分析 (Principal component analysis, PCA) 也曾经被用于前景提取^[52]; Hosaka 等^[53] 则利用支持向量机 (Support vector machine, SVM) 进行前景/背景分类; 而 Won 等^[54] 考虑到局部线性嵌入 (Locally linear embedding, LLE) 在把高维度数据映射到低维度时, 能保持点的邻近关系, 从而实现了一种抠像中应用 LLE 进行颜色估计的方法. 文献 [55] 中提出一种基于局部学习的方法, 构造出包含多次相似性权值计算的拉普拉斯矩阵, 精细地刻画数据局部几何结构, 从而求出更优的结果.

文献 [56] 则把抠像问题看成一个模式识别问题: 在学习阶段, 使用非负矩阵因式分解 (Non-negative matrix factorization, NMF) 构造表征前景及背景的基矩阵和系数矩阵; 在识别阶段, 根据未知像素的颜色值, 分解出原有的基矩阵与新的系数矩阵, 通过比较新的系数矩阵和原来的系数矩阵, 从而判断未知像素的不透明度 α . 算法中, 首先对未知像素 i , 分别选取两个距离最近的前景节点和背景节点及其 8 邻域像素的颜色值 (或者灰度值) 组成矩阵 V . 其中 V 共有四列, 取其中一个前景节点及其 8 邻域像素的颜色值组成第一列, 取另外一个前景节点及其 8 邻域像素的颜色值组成第二列. 对于后两列则使用同样的方法根据两个背景节点生成. 然后对 V 进行 NMF 分解, 即 $V \approx W \times H$. V 中的第 k 列数据可看做 W 的各行分别与 H 的第 k 列数据 H_k 的线性组合. 因此, 把 W 看做表征前景和背景的基矩阵, H 的前两列是对应前景特征的系数, 后两列对应背景特征的系数. 分解完成后, 即完成学习阶段, 开始识别阶段. 类似地, 取像素 i 及其 8 邻域像素的数据组成矩阵 V_{obj} , 然后使用之前计算的基矩阵 W 对 V_{obj} 进行分解, 得出新的系数矩阵 H_{obj} . 接着计算 H_{obj} 与 H 前后两列的欧几里得距离. 定义

$$d_{fg} = \min_{1 \leq i \leq 2} \|H_{obj} - H_i\|^2$$

$$d_{bg} = \min_{3 \leq i \leq 4} \|H_{obj} - H_i\|^2 \quad (15)$$

其中, d_{fg} , d_{bg} 分别表征像素 i 与前景/背景的距离. 最后根据 d_{fg} 与 d_{bg} 的比值确定像素属于前景或背

景。

上述基于机器学习的方法都是数字抠像的新尝试,与前三类方法不同,这类方法没有从求解抠像方程的角度出发,而是从机器学习或模式识别的视角看待抠像问题。通过一个学习过程,降低算法对图像特性的假设。然而,目前这种方法研究相对较少,上述方法也只是初步的尝试,例如没有对学习中使用的样本进行深入考虑, α 的估算也相对简单,缺少优化处理。

除了之前介绍的四类方法,在研究中,针对特殊的应用场景,也有其他的尝试。例如文献[57]使用图像的深度信息辅助抠像。一方面,在原有的颜色或灰度通道外,添加深度作为新的通道。另一方面,把 α 的梯度与深度图的梯度联系起来; Lin 等^[58]专门针对运动造成的模糊图像抠像问题进行了研究,通过图像梯度的统计信息估计物体的运动,从而提高抠像的准确度; Prabhu 等^[59]提出一种基于卡尔曼滤波器的方法对受损的图像进行抠像并恢复前景; 图像修复 (Inpainting) 技术也被用于抠像:由于目前的抠像方法一般通过邻近像素信息进行推导,没有考虑物体本身的几何形状,从而导致了一些抠像结果中原本平滑的线条出现了抖动。出于这种考虑,文献[60]使用变分修复 (Variational inpainting) 技术进行抠像。而文献[61]把未知区域看做受损或被物体遮挡,使用 Inpainting 技术估计其背景色; 文献[62]则认为,若已知将要合成的背景,抠像可以把目标设为,在特定的新背景中,增强最终合成结果在视觉上的效果。由此,提出的算法中把抠像与合成看作一个统一的过程; 文献[63]在抠像执行之前,使用区域增长 (Region growing) 以及分类技术对图像进行预分割。

3 分析和总结

3.1 测试结果分析

2000 年以后,针对自然图像的抠像算法逐渐增多,对抠像算法的评价方法也在逐步改进。最开始的评价方法中,仅凭人的视觉做主观判断,而现在量化评价逐渐被采用。2009 年, Rhemann 等 6 位学者,制定了一套较为全面和客观的测试集以及评价标准^[4],在近两年的论文中被广泛采纳。本文主要采用这套方案对目前的抠像算法进行分析。该测试集采用高质量的图像,其中高分辨率的图像在六百万像素左右,低分辨率的图像也超过 40 万像素。图像中包括了当前抠像技术中难以处理的场景,例如毛发、半透明物体、网状物体、纹理复杂的背景以及前景/背景颜色相似的区域等。Ground truth 由蓝屏抠像中的三角抠像法^[2]生成。对算法进行评价时,使用抠像结果与 Ground truth 进行对比。具体

地,包括四个标准: 均方误差 (Mean squared error, MSE)、绝对误差和 (Sum of absolute differences, SAD)、连通性误差以及梯度误差。梯度误差描述抠像结果与 Ground truth 中 α 值的梯度误差。与此类似,连通性误差描述抠像结果中,像素的连通度与 Ground truth 中对应像素的连通度间的误差。MSE 和 SAD 是常用的误差统计方法,而梯度和连通性误差是基于人的感官特点制定的。

根据这些测试集以及评价标准, Rhemann 等还建立了一个网上测试平台^[64]。由于目前的研究中,大部分算法以三分图或涂鸦方式作为用户的输入,且涂鸦结果可看作三分图的子集,支持涂鸦方式的算法也支持三分图,因此测试中使用三分图作为标准的输入方式,以保持评价的客观性。目前该平台中包含了截止至 2010 年主要抠像算法的测试结果。首先,针对这些测试结果,本文对主要的抠像算法进行定量和定性分析。其次,为了分析的全面性,也将对涂鸦方式和全自动化方式下的抠像结果进行对比。然后,对主要抠像算法的运行效率进行分析评价。

首先,本文引用了测试集中包含的 6 张测试图像,如图 3 所示。具体地,从本文第 2 节所分析的四类抠像技术中,各选取了一种抠像算法,对比其抠像结果的均方误差。这些算法分别为 Bayesian matting^[3], Closed-form matting^[16], Robust matting^[28] 以及 Learning based matting^[50]。测试结果与 Ground truth 的均方误差如图 4 所示。



图 3 测试图像

Fig. 3 Test images

可以看出,各种方法在不同的测试图像中,效果差别较大。对于 Bayesian matting,在图 3(b) 取得的效果相对较好,这是因为图 3(b) 中边缘比较简单,且前景物体中不存在孔状结构。采样得到的样本能反映未知像素的特性,比较有意义。而在测试图 3(c) 中表现最差,主要是因为图像的边缘及纹理比较复杂,前景物体中存在空隙,降低了样本的有效性。对颜色样本建立的模型不一定能对未知像素进行合适的描述。Closed-form matting 在前 5 张图像中表现比较稳定,然而在图 3(f) 中,则表现较差。这是因为算法假设前/背景颜色在局部区域内符合 Color-line

模型,生成的结果比较平滑,而图3(f)中的结构不符合这一假设。Robust matting因为在算法中融合了数据项和平滑项,在多幅图像中表现稳定。在前5幅图像中与Closed-form matting的结果大致相当。在图3(f)中,由于算法还结合了采集的样本信息,表现比Closed-form matting更好。Learning based matting作为新提出的方法,取得了与其他方法相当甚至更好的效果。特别是在包含半透明物体的图像(图3(e))中,效果比其他四种方法都好。说明算法在学习的过程中对模型进行更一般化的假设(例如非线性假设),可以增强算法在不同场景中的适应性。另一方面,这种方法对学习中使用的样本依赖性较大,由于实验中图3(f)对应的三分图前景区较小,而Learning based matting没有考虑采样等其他方式获得更多有效的学习样本,因此在图3(f)中没有取得好的效果。

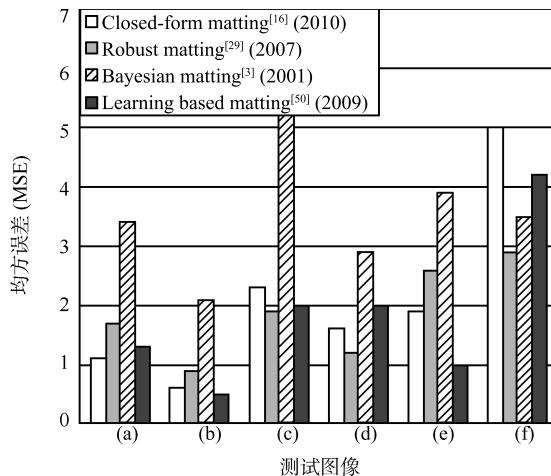


图4 多种算法抠像结果的均方误差

Fig. 4 MSE of matting results from various algorithms

此外,对更多的抠像算法进行测试分析,发现了一些总体来说效果最好的算法(如Shared matting^[30]、Improved color matting^[29]),这些算法往往不是单一地采用颜色采样或像素相似性的方法,而是融合了多种技术,并进一步对采样方式进行优化,或者构造了更有效的先验知识。图5对Improved color matting、Shared matting、Robust matting、Closed-form matting四种方法的抠像结果进行了比较。从图中明显地看出,前两种算法比后两种效果要好。事实上,Shared matting和Improved color matting算法中,首先进行前景/背景颜色采样,根据采集的样本求出初步的抠像结果,再结合Closed-form matting进行优化。此外,Improved color matting在采样时考虑了测地线距离,还在后续的 α 值估计时加入了“稀疏先验(Sparsity prior)”,这是根据颜色特征建立的,对估算出的 α

值进行偏移,使之等于0或1,以减少一些错误生成的模糊区。因此,可以认为,融合多种抠像技术、对采样技术进行优化、构造有效的先验知识是未来可能的发展方向。

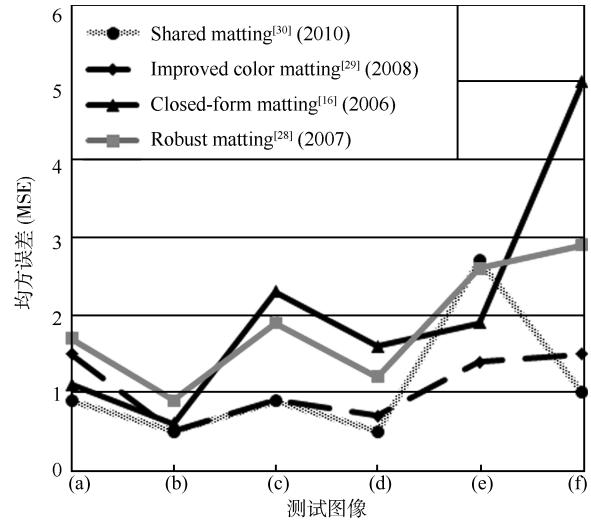


图5 多种算法抠像结果的均方误差比较

Fig. 5 Comparison of MSE of matting results from various algorithms

接着以测试图像中图3(c)的右下角为例,分析各种算法下抠像结果的差异。图6列举了六种算法的结果。可以看出, Bayesian matting^[3]在这种形状复杂的情况下表现不佳,出现了较多不连续的区域,而且由于背景与前景颜色相似,一部分背景被错误地抠选。Closed-form matting表现相对较好,然而,由于算法中的平滑假设,使得结果过于平滑,忽略了一些细节,而且由于没有颜色信息采集,有些深色的叶子被错误地判断为背景。Learning based 和 Robust matting表现相当,一些区域中仍然错误地把背景选中。Shared 和 Improved color matting则基本上消除了这个问题,但仍然有些瑕疵。

从上面的测试效果中,可以看出,在以三分图为算法输入的情况下,目前的算法已经取得了一定的进展。下面,分析其他用户交互方式下的抠像效果。从图7中可以看出,在前景/背景比较简单,且区分度较大的情况下,全自动化的Spectral matting算法可以生成较为理想的结果。而图8中则包含了对图3(a)的抠像结果。首先,图8(b)是Spectral matting^[20]的全自动抠像结果。可以看出,抠像结果中只提取了玩具的头发以及背景中的桥梁,没有选中玩具的身体,显然这个结果并不正确。这是由于缺乏用户输入,算法不能获得足够的语义信息,只能对视觉上特性接近的区域进行聚类。对于涂鸦的输入方式,图8(d)中展示了Closed-form matting的结果,由于输入图像的场景比较复杂,对于有限的涂鸦

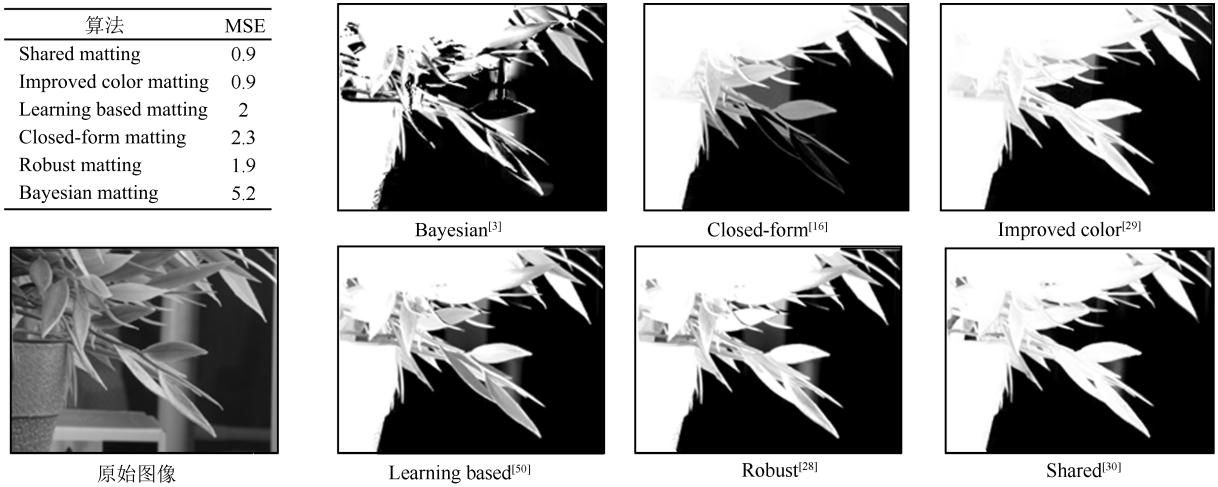


图 6 摹像结果比较

Fig. 6 Comparison of different matting results from various algorithms

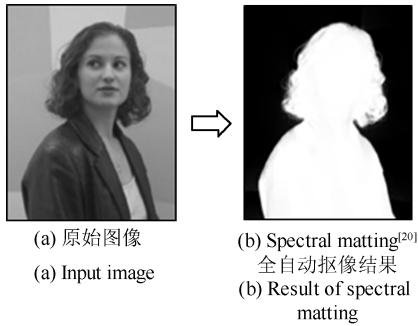


图 7 全自动抠像的结果

Fig. 7 Fully automatic matting result

标记, 算法不能从被标记的区域中获得未知区域的完整性. 而玩具头发的颜色和背景中桥梁的颜色比较接近, 因此部分桥梁区域被误选. 而玩具的手和衣服颜色区别较大, 手没有被完全选中. 图 8(e) 中的三分图则带有更精细的信息, 同样是 Closed-form matting 算法, 在三分图输入下获得了更好的结果(如图 8(f)). 由于算法基于 Color line 模型的假设, 以及桥梁颜色和头发颜色过于接近, 抹像结果中的右上角区域仍然不能精细地提取出头发丝. 根据图 7 和图 8 的结果, 可以看出, 相对于三分图的输入方式, 全自动方式的抹像更为简便, 在一些简单的场景中能得到理想的结果. 然而, 全自动或涂鸦方式下, 抹像算法更加缺乏语义信息, 对求解的不定问题进行限制, 因此求解过程变得更加困难, 对于复杂的场景, 抹像效果相对较差.

算法效率方面, 文献 [1] 中曾对一些有代表性的算法进行了测试. 测试程序运行于一台 CPU 主频为 3.0 GHz、内存大小为 4 GB 的工作站. 对于一张大小约 1000 像素 \times 700 像素的图像, 输入一张含有约 20% 未知像素的三分图, 算法运行时间如表 1 所示. 从表 1 可以看出, 这些算法均不能达到实时性的

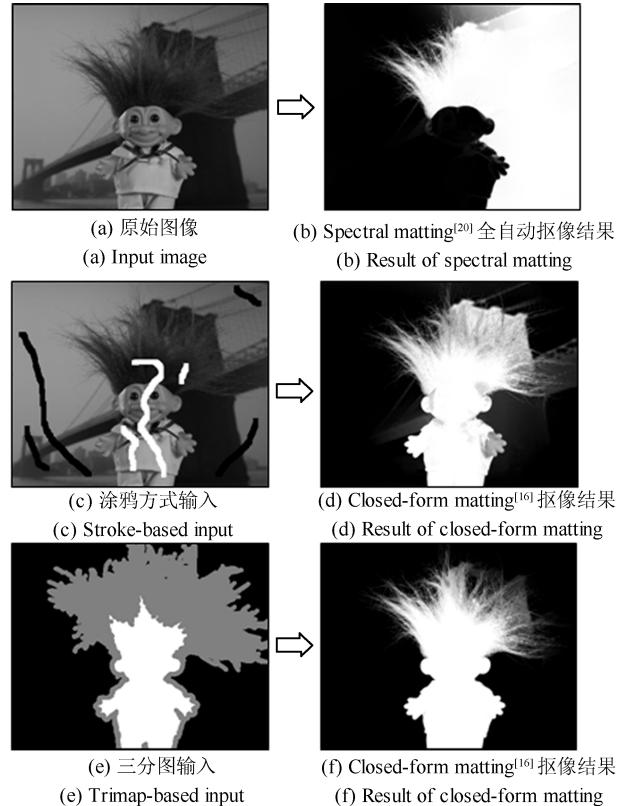


图 8 不同输入方式下的抹像结果

Fig. 8 Matting results of different interaction modes

要求. 其中, Bayesian matting 因为需要对每个未知像素的采集样本建立高斯混合模型, 所以消耗的时间较多. Iterative BP^[48] 则因为信任传播算法收敛较慢, 也带来较大的计算时间开销. Poisson matting 通过迭代的方法求解泊松方程, 效率也不高. Robust matting 由于结合了颜色采样技术, 以及 Closed-form matting 中的拉普拉斯矩阵进行求

解, 所以耗时较 Closed-form matting 多.

此后, 一些学术研究中, 通过使用图像处理器 (Graphics processing unit, GPU), 获得更高的抠像效率^[30, 65]. 这是因为, 虽然很多抠像算法中需要分别为每个未知像素进行采样计算, 或者最终归结为求解一个较大的线性系统, 但是这些求解过程一定程度上可以并行执行. 2010 年 Gastal 等提出的 Shared matting^[30] 通过 GPU 的并行加速后, 对于大小为 800 像素 \times 600 像素的图像, 大约只需 0.1 s 的处理时间. 如果使用 Shared matting 中的快速版本, 即除去基于 Closed-form matting 的后续优化步骤, 只使用基于采集样本的估计结果, 运行时间可进一步减少. 对于同样大小为 800 像素 \times 600 像素的图像, 只需 0.03 s. 虽然抠像效果略有下降, 但实现了实时的抠像.

表 1 各种算法的运行时间

Table 1 Running time of the algorithms

算法	运行时间 (s)
Bayesian matting ^[3] (2001)	721
Poisson matting ^[12] (2004)	26
Iterative BP matting ^[48] (2005)	507
Closed-form matting ^[16] (2006)	7
Robust matting ^[28] (2007)	14

3.2 总结和展望

根据上述的测试结果以及第 2 节的算法分析, 下面从精度以及效率方面对目前的抠像技术做如下总结.

1) 抠像精度方面, 通过图 4 的结果, 可以看出, 不同的技术在不同的图像中表现差别较大. 主要原因是不同的技术对图像进行了不同的假设. 一般地, 基于颜色采样的技术在前景/背景颜色分明、边界清晰的情况下表现较好. 而基于像素相似性的技术则能处理好颜色变化平滑及纹理稍微复杂的区域. 近年来很多方法融合了这两种技术, 因而取得了较好的结果. 然而, 对于网状、半透明或形状复杂的物体, 因为与通常算法中的假设不同, 或者不能够得到足够的有效样本信息进行求解, 仍然是抠像中的难点. 基于机器学习的方法对图像的假设较弱, 因此结果比较稳定, 但抠像结果受初始样本的限制仍然较大.

2) 求解效率方面, 由于各种算法的运行方式、用户交互程度各不相同, 很难给出一个统一标准进行比较. 不过, 对算法的分析可以看出, 一般地, 基于采样统计的技术进行样本聚类时, 需要较大的运算量; 基于相似性或能量函数的技术, 最后归结于最小化二次目标函数, 往往需要求解一个较大的

线性系统. 很多情况下, 这种线性系统的大小与图像大小成正比, 为求解效率带来很大的挑战. 因此, 一直有不少研究, 致力于实现能实时交互的数字抠像. 从 2004 年的 GrabCut^[19], 到 2007 年的 Soft scissor^[13], 还有 2010 年的 Large kernel matting^[39] 等, 速度方面已经取得一定的提高. 另一方面, 也有基于 GPU 的抠像方案^[30, 65–66], 得益于算法和硬件的可并行性, 这些方案在效率提升方面取得了进展, 大大地降低了求解时间. Shared matting^[30] 对于 800 像素 \times 600 像素大小的图片已经取得了实时的效果. 不过, 对于高清图像和视频的抠像, 求解效率还需要进一步提高.

数字抠像需要快速、高精度地完成, 同时为用户提供友好的交互方式. 目前抠像问题中还存在很多挑战性问题, 主要体现在以下四个方面:

1) 用户交互方面的研究. 从三分图到涂鸦式的输入, 抠像的交互方式已经得到了很大的简化. 然而, 不管是三分图还是涂鸦方式, 在应对某些场景时仍然显得不够直观, 例如特别纤细或存在孔状结构的物体. 同时, 对于一般的三分图和涂鸦方式, 用户的输入都独立于抠像算法的求解. 有时候用户的标记中含有许多冗余信息, 在一些关键的地方却缺少标记. 此外, 算法的计算效率、操作的复杂性等问题仍然影响着用户的输入和算法的可用性. 因此, 引导用户的输入、设计更好的交互流程或提高算法效率都有利于提高数字抠像的实用价值.

2) 采样技术的研究. 近年来, 数字抠像中采样技术得到不少关注, 而且已经取得了一定的效果^[30–31]. 因为数字抠像问题始终是一个病态的问题, 因此采集有效的样本是增加问题已知信息的重要渠道. 候选样本集的建立以及样本的选择都是采样技术的研究重点. 由于目前对候选样本的评价标准仍然局限于颜色空间距离和位置距离, 会导致一些原本相关性不大的样本被误选, 可以考虑添加纹理、几何形态等局部区域信息, 提高采样质量.

3) 先验知识方面的研究. 通过先验知识或假设, 为抠像问题构建有效的求解约束, 也是一个研究热点. 因为抠像问题在数学上是一个不定问题. 全自动的方式或者用户提供的有限标记, 都不能完整地对问题进行约束. 在一些纹理比较复杂的场景, 或者前景/背景色接近的区域, 算法经常无法正确地区分前景/背景. 这也是目前抠像研究中面临的主要问题之一. 然而, 在一些特定的应用背景下, 常常可以获得目标图像的一些先验知识或额外信息 (例如前景物体的颜色和形状的先验知识、图像对应的深度图 (Depth map) 等). 利用这些信息对问题进一步地约束, 是提高抠像效果的有效方法.

4) 基于机器学习的研究. 基于学习的方法可以减少对图像特性的假设, 通过样本训练建立前景不

透明度与图像颜色的关系, 将增大算法的健壮性, 适应多种类型的图像。然而, 如何进行特征提取, 提高学习的有效性值得进一步研究。此外, 如何通过有限的初始样本进行学习和推理, 目前也是一个挑战性的研究课题。

References

- 1 Wang J, Cohen M F. Image and video matting: a survey. *Foundations and Trends in Computer Graphics and Vision*, 2007, **3**(2): 97–175
- 2 Smith A R, Blinn J F. Blue screen matting. In: Proceedings of the 23rd ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). New Orleans, USA: ACM, 1996. 259–268
- 3 Chuang Y Y, Curless B, Salesin D H, Szeliski R. A Bayesian approach to digital matting. In: Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition. Kauai, USA: IEEE, 2001. 264–271
- 4 Rhemann C, Rother C, Wang J, Gelautz M, Kohli P, Rott P. A perceptually motivated online benchmark for image matting. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 1826–1833
- 5 Lin Sheng-You, Pan Rui-Fang, Du Hui, Shi Jiao-Ying. A survey on digital matting. *Journal of Computer-Aided Design and Computer Graphics*, 2007, **19**(4): 473–479
(林生佑, 潘瑞芳, 杜辉, 石教英. 数字抠图技术综述. 计算机辅助设计与图形学学报, 2007, **19**(4): 473–479)
- 6 Zongker D E, Werner D M, Curless B, Salesin D H. Environment matting and compositing. In: Proceedings of the 26th ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). Los Angeles, USA: ACM, 1999. 205–214
- 7 Yeung S K, Tang C K, Brown M S, Kang S B. Matting and compositing of transparent and refractive objects. *ACM Transactions on Graphics*, 2011, **30**(1): 1–13
- 8 Wu T P, Tang C K, Brown M S, Shum H Y. Natural shadow matting. *ACM Transactions on Graphics*, 2007, **26**(2): 1–30
- 9 Chuang Y Y, Goldman D B, Curless B, Salesin D H, Szeliski R. Shadow matting and compositing. *ACM Transactions on Graphics*, 2003, **22**(3): 494–500
- 10 Kim J, Kim J H, Joo S H, Choi B T, Jeong I K. Volume matting: object tracking based matting tool. In: Proceedings of the 2010 ACM Conference on Computer Graphics and Interactive Techniques in Asia (SIGGRAPH ASIA). Seoul, South Korea: ACM, 2010. Article 42
- 11 Xiao J J, Shah M. Accurate motion layer segmentation and matting. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 698–703
- 12 Sun J, Jia J Y, Tang C K, Shum H Y. Poisson matting. In: Proceedings of the 2004 ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). Los Angeles, USA: ACM, 2004. 315–321
- 13 Wang J, Agrawala M, Cohen M F. Soft scissors: an interactive tool for realtime high quality matting. *ACM Transactions on Graphics*, 2007, **26**(3): Article 9
- 14 Zheng Y J, Kambhamettu C, Yu J Y, Bauer T, Steiner K. FuzzyMatte: a computationally efficient scheme for interactive matting. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Alaska, USA: IEEE, 2008. 1–8
- 15 Guan Y, Chen W, Liang X, Ding Z A, Peng Q S. Easy matting: a stroke based approach for continuous image matting. *Computer Graphics Forum*, 2006, **25**(3): 567–576
- 16 Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **30**(2): 228–242
- 17 Rhemann C, Rother C, Rav-Acha A, Sharp T. High resolution matting via interactive trimap segmentation. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Alaska, USA: IEEE, 2008. 1–8
- 18 Chang H, Yang Q, Pan C H. An iterative Bayesian approach for digital matting. In: Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006. 122–125
- 19 Rother C, Kolmogorov V, Blake A. “GrabCut”: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 2004, **23**(3): 309–314
- 20 Levin A, Rav-Acha A, Lischinski D. Spectral matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(10): 1699–1712
- 21 Du W W, Urahama K. Image and video matting with membership propagation. In: Proceedings of the 8th Asian Conference on Computer Vision (ACCV). Tokyo, Japan: Springer-Verlag, 2007. 590–600
- 22 Sun J, Li Y, Kang S B, Shum H Y. Flash matting. *ACM Transactions on Graphics*, 2006, **25**(3): 772–778
- 23 Bleyer M, Gelautz M, Rother C, Rhemann C. A stereo approach that handles the matting problem via image warping. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 501–508
- 24 McGuire M, Matusik W, Pfister H, Hughes J F, Durand F. Defocus video matting. *ACM Transactions on Graphics*, 2005, **24**(3): 567–576
- 25 Berman A, Vlahos P, Dadourian A. Comprehensive Method for Removing from an Image the Background Surrounding a Selected Object. U. S. Patent 6134345, October 2000
- 26 Ruzon M A, Tomasi C. Alpha estimation in natural images. In: Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, USA: IEEE, 2000. 18–25
- 27 Juan O, Keriven R. Trimap segmentation for fast and user-friendly alpha matting. *Variational, Geometric, and Level Set Methods in Computer Vision*, 2005, **3752**: 186–197
- 28 Wang J, Cohen M F. Optimized color sampling for robust matting. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minnesota, USA: IEEE, 2007. 17–22
- 29 Rhemann C, Rother C, Gelautz M. Improving color modeling for alpha matting. In: Proceedings of British Machine Vision Conference. Leeds, UK: BMVA, 2008. 1155–1164
- 30 Gastal E S L, Oliveira M M. Shared sampling for real-time alpha matting. *Computer Graphics Forum*, 2010, **29**(2): 575–584

- 31 He K M, Rhemann C, Rother C, Tang X O, Sun J. A global sampling method for alpha matting. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 2049–2056
- 32 Lin S Y, Shi J Y. Fast natural image matting in perceptual color space. *Computers and Graphics*, 2005, **29**(3): 403–414
- 33 Cho S Y, Byun H. Adaptive color curve models for image matting. In: Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Turkey: IEEE, 2010. 225–228
- 34 Bai X, Sapiro G. A geodesic framework for fast interactive image and video segmentation and matting. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 35 Ning Q H, Wang W Q, Zhu C F, Qing L Y, Huang Q M. A fast approach for natural image matting using structure information. In: Proceedings of the 2007 IEEE International Conference on Multimedia and Expo. Beijing, China: IEEE, 2007. 1399–1402
- 36 Du Z L, Lin H, Qin X Y, Bao H J. Oriented poisson matting. In: Proceedings of the 2005 IEEE International Conference on Image Processing. Genoa, Italy: IEEE, 2005. 626–629
- 37 Price B L, Morse B, Cohen S. Geodesic graph cut for interactive image segmentation. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 13–18
- 38 Bai X, Sapiro G. Distancecut: Interactive segmentation and matting of images and videos. In: Proceedings of the 2007 IEEE International Conference on Image Processing. San Antonio, USA: IEEE, 2007. 249–252
- 39 He K M, Sun J, Tang X O. Fast matting using large kernel matting Laplacian matrices. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 2165–2172
- 40 Singaraju D, Vidal R. Interactive image matting for multiple layers. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Alaska, USA: IEEE, 2008. 1–7
- 41 Lee P, Wu Y. Nonlocal matting. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 2193–2200
- 42 Lee P G, Wu Y. L_1 matting. In: Proceedings of the 17th IEEE International Conference on Image Processing. Hong Kong, China: IEEE, 2010. 4665–4668
- 43 Singaraju D, Rother C, Rhemann C. New appearance models for natural image matting. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 659–666
- 44 Chen S F, Li Z G, Liu J Z, Tang X O. Image matting using linear optimization. In: Proceedings of the 15th ACM International Conference on Multimedia. New York, USA: ACM, 2007. 321–324.
- 45 Rhemann C, Rother C, Kohli P, Gelautz M. A spatially varying PSF-based prior for alpha matting. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 2149–2156
- 46 Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proceedings of 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 105–112
- 47 Li Y, Sun J, Tang C K, Shum H Y. Lazy snapping. *ACM Transactions on Graphics*, 2004, **23**(3): 303–308
- 48 Wang J, Cohen M F. An iterative optimization approach for unified image segmentation and matting. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005. 936–943
- 49 Park Y, Yoo S I. A convex image segmentation: extending graph cuts and closed-form matting. In: Proceedings of the 10th Asian Conference on Computer Vision. Queenstown, New Zealand: Springer-Verlag, 2010. 355–368
- 50 Zheng Y J, Kambhamettu C. Learning based digital matting. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 889–896
- 51 Draper N R, Smith H. *Applied Regression Analysis (Second edition)*. New York: John Wiley, 1981
- 52 Hillman P, Hannah J, Renshaw D. Alpha channel estimation in high resolution images and image sequences. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, USA: IEEE, 2001. 1063–1068.
- 53 Hosaka T, Kobayashi T, Otsu N. Image matting based on local color discrimination by SVM. *Pattern Recognition Letters*, 2009, **30**(14): 1253–1263
- 54 Won K H, Park S Y, Jung S K. Natural image matting based on neighbor embedding. *Computer Vision/Computer Graphics Collaboration Techniques*, 2007, **4418**: 449–460
- 55 Peng Hong-Jing, Chen Song-Can, Zhang Dao-Qiang. A local learning approach for natural image matting. *Journal of Software*, 2009, **20**(4): 834–844
(彭宏京, 陈松灿, 张道强. 一种基于局部学习的自然图像景物提取方法. 软件学报, 2009, **20**(4): 834–844)
- 56 Wang K, Zhang N N, Liu W X. Natural image matting with non-negative matrix factorization. In: Proceedings of the 2005 IEEE International Conference on Image Processing. Genoa, Italy: IEEE, 2005. 1186–1189
- 57 Pitie F, Kokaram A. Matting with a depth map. In: Proceedings of the 17th IEEE International Conference on Image Processing. Hong Kong, China: IEEE, 2010. 21–24
- 58 Lin H T, Tai Y W, Brown M S. Motion regularization for matting motion blurred objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **33**(11): 2329–2336
- 59 Prabhu S M, Rajagopalan A N. Natural matting for degraded pictures. *IEEE Transactions on Image Processing*, 2011, **20**(12): 3647–3653
- 60 Ni K Y, Sheshadri T, Chan T F. Matting through variational inpainting. In: Proceedings of the 9th IASTED International Conference on Signal and Image Processing. Honolulu, USA: ACTA, 2007. 179–184
- 61 Sarim M, Hilton A, Guillemaut J Y, Kim H. Non-parametric natural image matting. In: Proceedings of the 16th IEEE International Conference on Image Processing. Cairo, Egypt: IEEE, 2009. 3177–3180

- 62 Wang J, Cohen M F. Simultaneous matting and compositing. In: Proceedings of the 2007 ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). Minnesota, USA: IEEE, 2007. 1–8
- 63 Ruchanurucks M, Ogawara K, Ikeuchi K. Integrating region growing and classification for segmentation and matting. In: Proceedings of the 15th IEEE International Conference on Image Processing. San Diego, USA: IEEE, 2008. 593–596
- 64 Rhemann C, Rother C, Wang J, Gelautz M, Kohli P, Rott P. Alpha matting evaluation website [Online], available: <http://www.alphamatting.com>, May 16, 2012
- 65 Huang M C, Liu F, Wu E H. A GPU-based matting Laplacian solver for high resolution image matting. *The Visual Computer*, 2010, **26**(6–8): 943–950
- 66 Grady L, Schiwietz T, Aharon S, Westermann R. Random walks for interactive alpha-matting. In: Proceedings of the 2005 International Conference on Visualization, Imaging and Image Processing. Benidorm, Spain: ACTA, 2005. 423–429



张展鹏 中山大学信息科学与技术学院硕士研究生。中国科学院深圳先进技术研究院客座学生。主要研究方向为机器学习与计算机视觉。

E-mail: zhzhanp@gmail.com

(ZHANG Zhan-Peng) Master student at the School of Information Science and Technology, Sun Yat-Sen University and visiting student at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His re-

search interest covers machine learning and computer vision.)



朱青松 中国科学院深圳先进技术研究院医疗机器人与微创手术器械研究中心助理研究员。2010年获中国科学技术大学硕士学位。主要研究方向为机器学习与机器人视觉技术基础研究。本文通信作者。E-mail: qs.zhu@siat.ac.cn

(ZHU Qing-Song) Assistant professor in Research Centre for Medical Robotics and Minimally Invasive Surgical Devices, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He received his master degree from University of Science and Technology of China in 2010. His research interest covers computer vision and robotics. Corresponding author of this paper.)



谢耀钦 中国科学院深圳先进技术研究院研究员。2002年获清华大学博士学位。主要研究方向为图像引导放射治疗, 医学影像处理和分析。

E-mail: yq.xie@siat.ac.cn

(XIE Yao-Qin) Professor at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. He received his Ph. D. degree from Tsinghua University in 2002. His research interest covers image guided radiation therapy, medical images processing and analysis.)