基于前景与透明度联合求解的多层次仿射类抠像与合成方法

姚桂林^{1,2}, 张必英^{1,2}, 张艳荣^{1,2}, 苏晓东^{1,2*}

1. 哈尔滨商业大学计算机与信息工程学院, 哈尔滨 150028

2. 黑龙江省电子商务与信息处理重点实验室, 哈尔滨 150028

* 通信作者. E-mail: suxd@hrbcu.edu.cn

收稿日期: 2020-05-06; 修回日期: 2020-07-24; 接受日期: 2020-09-14; 网络出版日期: 2021-10-12

黑龙江省自然科学基金项目 (批准号: F2018021) 和哈尔滨商业大学校级科研项目 (批准号: 18XN021, 2020CX39) 资助

摘要 数字图像抠像与合成问题是虚拟现实中的经典问题. 抠像问题需要根据抠像公式, 从输入图像准确地提取出前景物体的透明度 (α) 和前景物体的颜色值, 而合成问题仅使用抠像结果与新背景图像进行结合, 使其成为新的合成图像. 然而, 目前很多抠像算法仅局限于抠像部分的 α 值的计算, 忽略了用于后续合成步骤的前景颜色的计算. 这实际上仅限于理论研究, 脱离了抠像问题的合成应用层面. 同时, 其他一些兼顾计算前景颜色算法的效果仍然不理想, 而且也仍然缺乏合成图像数据集和对合成图像的比较. 本文充分利用仿射类方法中搜索范围灵活可变的优势, 采用实体像素与混合像素进行分级处理的 3 层次结构, 联合求解出最终的 α 值和前景颜色. 其中, 每个层次均利用前向层次的先验, 结合仿射类方法的像素相关性以及颜色拟合性约束, 采用闭合形式进行求解. 在实验中, 本文不仅针对 α 层面进行比较, 而且在数字抠像领域中首次引入针对合成图像层面的比较. 在合成层面中, 本文针对输入图像及各型号的三分模板 (Trimap), 均引入了白色背景和相应的自然图像背景. 同时各方法的真实的 α 值和前景颜色不仅与这些背景进行了常规合成, 而且也进行了一种单纯 α 层面的合成. 实验结果表明, 在数值评价结果和视觉效果方面, 本文方法的最终 α 结果与上述各种合成效果均优于目前已有的抠像与合成方法及一些新生成的方法.

关键词 图像抠像, 图像合成, 仿射类抠像, 闭合形式解, 实体像素与混合像素

1 引言

数字图像抠像与合成问题是虚拟现实领域中的经典问题, 其主要目标将提取前景目标信息, 并将其合成到新背景中. 在数学描述中, 图像中的某点 i 的颜色 I_i 是它的前景颜色 F_i 与背景颜色 B_i 的

引用格式: 姚桂林, 张必英, 张艳荣, 等. 基于前景与透明度联合求解的多层次仿射类抠像与合成方法. 中国科学: 信息科学, 2021, 51: 1658–1674, doi: 10.1360/SSI-2020-0121
Yao G L, Zhang B Y, Zhang Y R, et al. A multi-level affinity-based matting and compositing method based on a joint solution of foreground and transparency (in Chinese). Sci Sin Inform, 2021, 51: 1658–1674, doi: 10.1360/SSI-2020-0121

线性组合:

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

其中 α_i 为 i 点的透明度因子且 $\alpha_i \in [0, 1]$. $\alpha_i = 1$ 或 $\alpha_i = 0$ 表示实体前景像素与实体背景像素, 而 $0 < \alpha_i < 1$ 的点称作混合像素, 通常出现在前景物体的边缘处及半透明物体内部. 数字抠像步骤可描述为给出 3 通道的合成图像 I , 根据抠像公式 (1), 求解出 1 个通道的 α 和 3 个通道的 F 与 B , 形成 $\{F, B, \alpha\}$ 求解序列. 而合成步骤则是抽取抠像步骤求得的 $\{F, \alpha\}$ 值, 并结合新背景 B' 与抠像公式 (1), 计算出新的合成图像 I' . 显然, 抠像步骤的求解是一个欠约束问题, 同时也是难点问题与关键问题. 同时, 合成步骤虽然比抠像步骤简单很多, 但却是抠像问题的终极目标: 单纯进行的抠像步骤也仅为理论层面的分析, 与合成问题合二为一才能达成实际应用的目的.

由于抠像步骤是欠约束问题, 因此需要用户手工输入一个三分模板, 记为 Trimap, 用来粗略地标注一部分实体前景区域 Ω_F 与实体背景区域 Ω_B , 剩余像素为需要后续计算的像素区域 Ω_U , 包括全体混合像素与部分未标注的实体像素. 有些时候为了简化用户输入, Trimap 退化为较为粗略的线条, 此时标注的实体前景和背景像素较少, 而需要计算的剩余像素较多.

由抠像公式可以看出, 前景 F 反映出每个像素的色彩, 而 α 通道则反映出前景色彩的强与弱 (或“浓与淡”) 以及前景物体的大致轮廓 (体现在 α 接近于 0 时轮廓逐渐消失). 传统观念认为, 对 α 的求解是抠像步骤的核心目标与难点, 也体现了抠像算法成败的决定因素. 另外, 由于前景 F 与背景 B 相比 α 要平缓很多, 各类抠像算法大多忽略了对 F 和 B 的计算, 其中甚至包括目前主流的基于深度神经网络的抠像类方法. 然而, 单独 α 的计算仅局限于理论研究层面, 并无法进行合成. 某些方法甚至引入了初始图像 I 代替 F 进行合成, 但这会产生“背景色溢出”现象, 将在后续实验中详细阐述. 事实上, F 的计算是后续合成步骤的必要条件之一, 只有同时计算 F 才是将抠像问题引领至实际应用的层面. 另一方面, α 通道仅包含 1 个未知量, 而 F 包含了 3 个未知量, 纵使 F 的颜色分布较 α 平缓, 但从折衷的角度来说, 对 F 求解的复杂程度并不低于求解 α .

前文提到, 近期的深度神经网络类抠像方法^[1~3] 仅局限于 α 计算, 而没有出现同时求解前景颜色的网络架构, 这使得它们尚难以应用至合成层面. 在传统方法中, 计算 α 的方式包括采样类与仿射类方法, 相应地, 在已出现的计算 F 的方法中, 计算 F 的方式也同样对应采样类与仿射类方法. 另外也包括一些其他特殊 α 计算方法, 由于它们需要通过采样类方法求得 F , 因此将其与采样类方法进行合并. 以下着重回顾同时计算 $\{F, B, \alpha\}$ 的两类方法.

第 1 类为采样类方法. 该类计算 α 方法的综述详见文献 [4]. 对于利用采样类方法计算 F (与 B) 的方法, 可以概括为隐式计算与显式计算 2 种. Robust^[5], Comprehensive^[6], WCT^[7], KL-D-Sparse^[8] 等采样类抠像方法是一类隐性计算 F 的方法, 它们在计算 α 时并未明确提及 F 的计算, 但在利用采样方法计算初级 α 时, 可以同步记录并还原最优 α 所对应的 F (与 B) 的样本值. 然而, 该类方法均需要利用仿射类方法进行后处理, 此时无法计算通过加入仿射项计算 α 后的 F 的值. Shared 算法^[9] 通过显式方法明确计算出 F (与 B) 的值, 它同样在利用采样方法计算 α 时, 保留了当前点最优的 F 与 B 值. 然而该方法可以不采用仿射类方法进行后处理, 而是可选用一个 α 平滑步骤同步求出最终的 F (与 B). Bayesian 方法^[10] 采用贝叶斯框架同时求解 α 及 F 与 B , 鉴于其采用局部采样方式, 可以将它看作大致与 Robust 方法相近的显式求解方法.

Geodesic 方法^[11] 由于其对特殊的对全体 α 同时求解的方式, 在求解 α 过程中无法同步记录前景与背景的相关样本. 然而, 该方法之后采取了一种采样类方式, 在已知前景与背景边缘处采集若干局部样本, 最优 F 样本与 B 样本在基于已求得的 α 和输入的合成图像 I 的基础上, 使得抠像公式 (1) 拟合

误差最小. 事实上, 此种方法同样适用于基于全局采样类的 Global^[12]、稀疏编码方法 Sparse-Coded^[13], 以及仿射类方法等在求解 α 的过程中无法记录相应样本的方法, 它的具体表现形式及应用将在后续的实验中讨论. 与采样类方法求解 α 的结果类似, 采样类求解 F 的方式也容易在前景 (和背景) 结果中产生不连续性的问题, 同时对实体像素的求解也不够理想.

第 2 类为仿射类方法. 该类计算 α 方法的综述详见文献 [14]. 该类方法的优点是强调了 α 的空间一致性, 使其在平滑性与视觉效果上要优于采样类方法. 仿射类方法计算 α 的实质是针对邻域像素的线性组合, 各方法的主要区别是对于邻域像素的搜索范围以及权重的计算方式. Closed form 方法^[15]采用最近邻搜索方式及 Matting Laplacian 权重, nonlocal 方法^[16]与 KNN 方法^[17]分别采用较近距离与中远距离搜索范围及简单权重的方式, LNSP 方法^[18]采用 closed form 与 KNN 类方法权重的简单求和, CCM^[19]方法采用 KNN 搜索方法结合 Matting Laplacian 权重的方法, information flow 方法^[20]在 LNSP 方法的基础上进一步引入了对已知像素的直接搜索方法, three-layer^[21]方法采用了基于 3 种 KNN 距离下的搜索与 Matting Laplacian 权重相结合的方法.

与采样类方法不同, 仿射类方法由于同时求出全部像素的 α , 因此无法记录求解过程中各点的 F (与 B) 值. 若需对 F (与 B) 求解, 则需要设计新的显式求解方法. Closed form 方法的求解方式与 random walk 方法^[22]求解 α 的思想类似, 在求出 α 后, 将当前点的 F (与 B) 值视为其 8 邻域 F (与 B) 的简单线性组合, 并以当前点与 8 邻域的 α 差异作为权重, 再利用 3 次闭合形式求解的方式, 分别求出 3 个通道的 F 值和 B 值. 该方法虽然不适用于求解具有空间变化特性的 α , 但由于 F (与 B) 的空间分布通常较 α 平缓, 因此该方法在一般的硬边界或相对较窄的半透明毛发处对 F (与 B) 的求解效果比较好. 然而, 该方法中 α 的差异并无法准确代表 F (或 B) 的差异, 而且由于采用最近邻的方式, 在 F (或 B) 的空间颜色变化较剧烈的情况下影响了较远距离样本的选取. 除上述方法外, nonlocal, KNN, LNSP, CCM, information flow 等仿射类 α 求解方法均未设计针对 F (或 B) 进行单独求解的方法.

本文求解 F (与 B) 的方法与 closed form 中的方法类似, 均采用线性组合与颜色拟合相结合的方式, 然而本文具有以下 3 点优势. 第一, 本文采用搜索范围更远的 KNN 方式取代最近邻方法, 以期获取空间距离更远的样本; 第二, 本文基于我们之前 Trimap 预处理综述^[23]中分步求解的思想, 设计出相应地由实体像素向混合像素递进求解的 3 层次结构, 并最终同时求出 α 与 F ; 第三, 本文借鉴了 late fusion 方法^[2]的思想, 并不局限于 α 本身的计算, 而是同时考虑对 α 反置的计算 (即背景概率).

2 基于仿射类的多层次求解方法

本文利用仿射类方法结合先验与拟合性约束的方式, 分 3 个层次求解 α 和前景颜色. 第 1 层利用基于 KNN 方式的远距离搜索或者近距离搜索的仿射类方法, 针对实体像素计算初级 α^{pre} 值, 并为后续步骤提供实体像素的先验值; 第 2 层结合上述先验, 引入前景和背景的直接搜索与间接搜索以及拟合性约束, 针对混合像素计算初级前景 F^{pre} 与背景 B^{pre} , 并为后续步骤提供混合像素的 α 与前景先验值; 第 3 层结合上述所有先验, 引入联合性 Matting Laplacian 抠像矩阵与拟合性约束, 同步求出最终的 α^* 与前景值 F , 并与新背景 B' 合成图像 I' . 整体抠像与合成流程如图 1 所示.

2.1 α 预计算

在文献 [23] 中强调, 实体像素计算步骤应该与混合像素相区分, 而且仿射类方法强调空间相关性与其搜索范围广的特点表明了该类方法有利于计算实体像素. 上述特点对于前景颜色 F 的计算同样

图 1 (网络版彩图) 本文的 3 层次抠像框架以及最终合成步骤

Figure 1 (Color online) The 3-level matting framework and the final compositing step included in this paper

有效. 事实上, 由于 $\alpha = 0$ 或 $\alpha = 1$ 同时表明 $F = 0$ 或 $F = I$, 即对实体像素 α 的求解也同时求出了 F , 因此并不需要独立的步骤预计算实体 F (或 B). 本节采用仿射类方式求解初始 α 并作为预划分的依据. 与硬性划分方法不同的是, 本文并非完全将 $\alpha = 0$ 或 $\alpha = 1$ 的结果判为后续步骤的已知前景或背景, 而是使这些点为后续计算提供实体像素与混合像素的 $\{F, B, \alpha\}$ 的先验值. 由于 Nonlocal 与 KNN 类搜索范围较广的方法具备优异的实体像素划分性能, 因此本小节采用该类方法计算一个初始 α 值, 对实体像素进行预计算.

KNN 类求解 α 方式的主要思想是当前未知点 i 是若干邻域 j 的线性组合. 根据方法 [20], KNN 类间接搜索的抠像方式可以依靠针对搜索不同距离的空间邻域, 调节 α 对实体像素的计算性能. 本文采用 FLANN 搜索方式下的 HSV 色彩空间进行初始抠像, 该搜索的源区域与目标区域的映射为 $\Omega_U \rightarrow \Omega_U \cup \Omega_F \cup \Omega_B$. 令每个像素 i 的特征为 $X_i = (C_{i1}, C_{i2}, C_{i3}, w_i^s x_i, w_i^s y_i)$, 其中前 3 个分量为归一化的色彩分量, 后 2 个分量 x_i 与 y_i 为归一化后的横坐标与纵坐标, w_i^s 为距离因子, 其值越大表明搜索范围越近. 在实践中, 若前景与背景颜色重合程度低, 则可以采用远距离搜索方式寻找更多样本, 否则需采用近距离搜索来消除样本间的高重合性造成的 α 求解误差.

本文首先计算每个前景与每个背景之间的颜色差异, 根据该差异分为以下 2 个搜索分支. (a) 若该差异大于某个阈值 $T_c = 0.1$, 则采用包含远距离搜索方式的 three-layer^[21] 方法. 大体上, 该方法采用在 RGB 色彩空间 (即 X_i 的前 3 个分量为归一化的 RGB 颜色) 下 w_i^s 分别取 1, 0.1, 0.01 下的 3 个联合搜索距离. 在这 3 个搜索范围下分别搜集 12, 3, 7 个最近邻样本 j . (b) 若该差异小于阈值 T_c , 则采用相对近距离的搜索方式. 鉴于 HSV 色彩空间对实体像素的判别相对较好, 这里我们将 X_i 的前 3 个分量置为 HSV 空间的各通道, 角度分量 H_i 也需要归一化至 $[0, 1]$ 区间, 权重 w_i^s 分为 0.5 与 0.1 两级, 这里去掉了 0.01 这一很大的搜索范围. 在这 2 个搜索范围下分别搜集 10 个与 2 个最近邻样本 j .

于是未知点 i 与样本 j 之间的权重为

$$w_{ij} = 1 - \frac{\sum_{c \leq 3} |X_i^c - X_j^c|}{3}, \quad (2)$$

其中 X_i^c 为特征 X_i 的第 c 个通道. 通过求解如下稀疏线性系统可求得该步骤的 α^{pre} :

$$(M^{\text{pre}} + \gamma D_\alpha) \alpha^{\text{pre}} = \gamma b_\alpha, \quad (3)$$

图 2 (网络版彩图) α^{pre} 的预计算及比较. (a) 输入图像与 Trimap; (b) 真实值 α^{true} ; (c) 采用 three-layer 方法计算的 α^{pre} ; (d) 采用 HSV 色彩空间计算的 α^{pre} , 其中蓝色、橙色、绿色分别表示扩展实体前景区域 $\Omega_{F'}$, 扩展实体背景区域 $\Omega_{B'}$, 剩余混合区域, 第 1 行与第 2 行分别表示前景与背景颜色差异较大与较小的例子

Figure 2 (Color online) Precomputation and comparison of α^{pre} . (a) Input image and trimap; (b) true value α^{true} ; (c) α^{pre} computed by three-layer method; (d) α^{pre} computed through the HSV color space, where blue, orange and green pixels represent the extended opaque foreground region $\Omega_{F'}$, extended background region $\Omega_{B'}$, and remaining mixed region respectively

其中 M^{pre} 为第 i 行对应样本 j 的权重为 w_{ij} 的稀疏矩阵, 其对角线为每行权重之和的负值; D_α 为对角阵, Trimap 对应的已知区域像素 ($\Omega_F \cup \Omega_B$) 为 1, 其余为 0; b_α 为列向量, Ω_F 为 1, 其余为 0; γ 为一个较大的值. 注意, 式 (3) 中的 γD_α 与 γb_α 项在后续两小节中同样用到, 为了简化, 在后续求解中省略了这两个必要约束, 而仅仅关注于未知区域 Ω_U 内的约束.

上述系统在求解后, 可将 α^{pre} 在 Ω_U 内扩展后的实体前景与实体背景区域记为 $\Omega_{F'}$, $\Omega_{B'}$, 表述为

$$\begin{cases} \Omega_{F'} = \Omega_U \cap \{i | \alpha_i^{\text{pre}} > 0.8\}, \\ \Omega_{B'} = \Omega_U \cap \{i | \alpha_i^{\text{pre}} < 0.2\}. \end{cases} \quad (4)$$

这里将第 1 节的两个实体区域进行了适当扩展, 是因为在 α 接近于 1 与 0 的区域也具备与原实体区域近似的特征, 而且两个区域并不更新原有的 Ω_F , Ω_B 与 Ω_U 等区域, 而是仅为后续计算提供实体像素的先验值. 这两个区域在后续计算中也需要用到, 具体实例如图 2 所示. 这 2 个实例表明在差异较大、重合程度不高时, 采用 three-layer 方法较好, 而基于 HSV 空间容易将混合像素误判为实体区域, 而在前景与背景差异较小时采用后者较好. 需要注意的是在该步骤中, 我们只关注实体区域而不关注剩余混合像素, 因此较少的误判别会有助于后续计算.

2.2 前景与背景的初级计算

该步骤根据上节的初始 α^{pre} , 计算每个未知点的初始前景与背景, 并后续步骤提供混合像素的先验值. 因为前景与背景均存在 3 个通道, 需要对每个通道单独计算. 求解第 c 个通道的 F^{c*} 与 B^{c*} 需要最小化如下函数:

$$\{F^{c*}, B^{c*}\} = \arg \min_{F^c, B^c} (V_1(F^c, B^c) + \beta_2 V_2(F^c, B^c) + \beta_3 V_3(F^c, B^c)). \quad (5)$$

上式分为 3 个函数 V_1, V_2, V_3 , 第 1 项的权重体现在函数内部, 后两项的权重为 β_2 与 β_3 , 以下逐一阐述.

式 (5) 的第 1 部分 V_1 为线性组合求解部分, 是该步骤仿射类方法的主体. 与仿射类 α 的求解方式类似, 该项体现了前景和背景所表现出的空间连续性, 表述如下:

$$V_1(F^c, B^c) = \sum_{i \in \Omega_U} \left(\left| W_i^F F_i^c - \sum_j w_{ij}^F F_j^c \right| + \left| W_i^B B_i^c - \sum_j w_{ij}^B B_j^c \right| \right), \quad (6)$$

这里的未知量 F_i^c 和 B_i^c 表示未知点 i 点第 c 个通道的前景和背景颜色, $W_i^F = \sum_j w_{ij}^F$, $W_i^B = \sum_j w_{ij}^B$. 其中, 点 i 的各前景样本 j 对应的权重 w_{ij}^F 分为直接样本权重与间接样本权重, 背景同理. 以下对这 2 个权重的计算做出详细的阐述.

传统的仿射类 α 求解方式中的间接搜索的方式无法体现这种前景与背景之间的插值方式. 因此对于前景和背景的搜索, 这里采用直接搜索与间接搜索相结合的方式, 将通过直接搜索方式获取的样本称为直接前景/背景样本, 间接搜索方式获取的样本称为间接前景/背景样本. 直接样本与间接样本的示例如图 3(a) 所示.

直接前景与直接背景样本的选取采用 FLANN 的搜索方式, 其源区域与目标区域的映射分别为 $\Omega_U \rightarrow \Phi_F$ 与 $\Omega_U \rightarrow \Phi_B$, 其中 Φ_F 与 Φ_B 分别为 Ω_F 与 Ω_B 区域中靠近未知区域处的 5 个像素宽度的 2 个已知区域. 在该搜索模式下, 对每个未知点 i 分别选取 10 个直接前景样本 F_j^{dir} 和直接背景样本 B_k^{dir} (这里的样本是 3 通道的向量), $i, k = 1, \dots, 10$. 因为近处的样本最能够反映出前景与背景真实颜色, 因此这里的特征定义为 $X = (w_c R, w_c G, w_c B, x, y)$, $w_c = 0.1$ 为颜色因子, 作用是尽量获取空间位置距离当前未知点较近已知样本, 去除远处的干扰性样本. 对于未知点 i , 首先计算它与样本 F_j^{dir} 和样本 B_k^{dir} 的拟合误差 D_{ijk} 为

$$D_{ijk} = \|I - \alpha_i^{\text{pre}} F_j^{\text{dir}} - (1 - \alpha_i^{\text{pre}}) B_k^{\text{dir}}\|. \quad (7)$$

之后, 定义每个直接前景样本 j 的拟合误差 $D_{ij}^F = \min_k (D_{ijk})$, 于是直接前景样本 j 的权重为

$$w_{ij}^{\text{Fdir}} = \exp((D_{ij}^F - \min_j (D_{ij}^F))^2 / \sigma_1^2), \quad (8)$$

其中 $\sigma_1^2 = 0.005$.

间接前景与间接背景样本的选取仍然采用 FLANN 的搜索方式, 这里源区域与目标区域的映射分别为 $\Omega_U \rightarrow \Omega_U \cup \Omega_F$ 与 $\Omega_U \rightarrow \Omega_U \cup \Omega_B$, 由于已知背景区域 Ω_B 不存在任何前景信息, 因此在搜索前景间接样本时需要除去 Ω_B , 背景同理. 实践表明, 采用最近邻方式计算前景与背景颜色的 closed form 方法会使得结果较为平滑、不宜受到前景与背景间颜色重叠影响, 而较大的搜索范围会使得混合点的前景与背景颜色计算不准确, 于是这里的搜索特征为 $X = (R, G, B, w_s x, w_s y)$ 、距离分量的权重为 $w_s^s = 200$, 使其接近于 closed form 方法中相邻样本的选取, 并分别搜索每个未知像素的 10 个间接前景样本 F_j^{ind} 和直接背景样本 B_k^{ind} , 样本权重 w_{ij}^{Find} 与 w_{ij}^{Bind} 均按式 (2) 计算, 此时特征 X_i 的前 3 个特征为 RGB 分量.

此外, 根据未知点距离已知前景区域的远近程度, 对每个未知点 i 与间接前景样本和直接前景样本定义附加权重 v_i^{Find} 与 v_i^{Fdir} (背景同理), 并按下式分配它们的权重:

$$\begin{cases} v_i^{\text{Find}} = \min(D_s(i, F) / D_t, 1), \\ v_i^{\text{Fdir}} = 1 - v_i^{\text{Find}}, \end{cases} \quad (9)$$

图 3 (网络版彩图) (a) 白色未知点的直接前景与背景样本点 (浅蓝色与黑色圆点) 和间接前景与背景样本点 (方块与三角); (b) α^{true} 的混合区域 $\Omega_{U'}$; (c) α^{mix} 的混合区域 $\Omega_{U'}$; (d) 信念值 f

Figure 3 (Color online) Direct search, indirect search and mixed α prior. (a) Direct foreground and background sample points (light blue and black dots) and indirect foreground and background sample points (squares and triangles) for white unknown point; (b) mixed region $\Omega_{U'}$ in α^{true} ; (c) mixed region $\Omega_{U'}$ in α^{mix} ; (d) confidence f

其中 $D_s(i, F)$ 为点 i 与 Ω_F 的最短空间距离, $D_t = 50$. 上式表明, 若未知点与前景边界距离较远, 则直接样本并不够可靠, 此时倾向于间接前景样本的选取; 反之若距离很近, 则倾向于直接前景样本. 最后, 对于式 (6) 中未知点 i 的每个样本 j , 若 j 为直接样本, 该样本的最终权重 $w_{ij}^F = w_{ij}^{\text{Fdir}} \cdot v_i^{\text{Fdir}}$, j 为间接样本时的权重同理可得. 同样方式可求得式 (6) 中各背景样本的权重 w_{ij}^B . 注意, 该权重对 3 个通道均适用, 这样在求解所有通道时可仅计算一次.

式 (5) 的第 2 部分 V_2 为第 1 步的 α^{pre} 提供的实体前景与实体背景的先验, 表述为

$$V_2(F^c, B^c) = \sum_{i \in \Omega_{F'}} (|F_i^c - I_i^c|) + \sum_{i \in \Omega_{B'}} (|B_i^c - I_i^c|), \quad (10)$$

其中, 这里的已知量 I_i^c 与未知量 F_i^c 与 B_i^c 均为第 c 个通道的颜色值, 由于此时混合点的前景与背景颜色尚未求出, 因此区域 $\Omega_{F'}$ 与 $\Omega_{B'}$ 由式 (4) 定义并将阈值分别取 1 与 0, 该部分权重为 $\beta_2 = 10$.

式 (5) 的第 3 部分为根据先验 α^{pre} 值、需求解的前景 F_i 和背景 B_i , 以及抠像公式 (1), 做出拟合性平衡, 表述为

$$V_3(F^c, B^c) = \sum_{i \in \Omega_U} |I_i^c - \alpha_i^{\text{pre}} F_i^c - (1 - \alpha_i^{\text{pre}}) B_i^c|, \quad (11)$$

同样, 这里的已知量 I_i^c 与未知量 F_i^c 与 B_i^c 均为第 c 个通道的颜色值, 与式 (7) 中的已知样本向量不同. 权重 $\beta_3 = 10$.

最后, 将 3 个通道的 F^{c*} 与 B^{c*} 的结果图像化, 并进行通道合并, 记为 3 通道图像的 F^{pre} 与 B^{pre} , 它们的实例如图 1 的第 2 层步骤所示. 同时, F^{pre} 与 B^{pre} 不仅能为最终的 F 提供混合像素先验, 也能提供一个 α 的混合像素先验 α^{mix} , 计算为

$$\alpha^{\text{mix}} = \frac{(I - B^{\text{pre}}) \cdot (F^{\text{pre}} - B^{\text{pre}})}{\|F^{\text{pre}} - B^{\text{pre}}\|^2}, \quad (12)$$

这里是逐点计算方式. 因其严格按抠像公式 (1) 求解, 因此较 α_i^{pre} 对混合像素的计算要精确, α^{true} 与 α^{mix} 计算混合点的实例如图 3(b) 和 (c) 所示.

2.3 α 与前景的优化

2.1 小节中, 由于 α_i^{pre} 针对实体像素, 因此对混合像素的求解并不够精确, 而且 2.2 小节中前景与背景的结果仅为 α_i^{pre} 的初级结果, 同时, α 与 F 并没有同步求出, 这也会为求解带来一定的误差, 因此上述结果均不作为最终结果. 本小节利用 α_i^{pre} 的实体像素先验, 并利用 F^{pre} 与 B^{pre} 及 α^{mix} , 作为

混合像素的先验, 结合仿射类方法以及抠像公式的拟合性平衡, 同时对 α 与 F 进行最优化求解. 最终的 $\tilde{\alpha}^*$ 与 F^* 通过最小化如下函数求得:

$$\{\tilde{\alpha}^*, F^*\} = \arg \min_{\tilde{\alpha}, F} (O_1(\tilde{\alpha}, F) + O_2(\tilde{\alpha}, F) + \theta O_3(\tilde{\alpha}, F)), \quad (13)$$

这里的 $\tilde{\alpha}^* = 1 - \alpha^*$ 为背景概率, α^* 为需要最终计算的 α 值. 优化 $\tilde{\alpha}$ 而不是 α 的目的不仅借鉴于 late fusion 方法丰富求解模式, 而且也便于采用仿射类方法优化函数 O_3 . 上式共分为 3 个函数 O_1, O_2, O_3 , 前 2 项的权重体现在各函数内部, 第 3 项的整体权重为 θ , 以下逐一阐述.

式 (13) 的第 1 部分 O_1 是线性组合部分及仿射类步骤的主体, 体现了 $\tilde{\alpha}$ 与前景颜色对于混合像素的空间连续性, 表述为

$$O_1(\tilde{\alpha}, F) = \sum_{i \in \Omega_U} \left(\left| W_i^{\tilde{\alpha}} \tilde{\alpha}_i - \sum_j w_{ij}^{\tilde{\alpha}} \tilde{\alpha}_j \right| + \left| W_i^F F_i^c - \sum_j w_{ij}^F F_j^c \right| \right). \quad (14)$$

由于之前的 α_i^{pre} 着重计算了 α 中的实体像素, 因此 O_1 的前半部分特别针对混合像素. 根据方法 [24], 仿射类中的混合像素计算需要采用利用抠像公式推导而得的 Matting Laplacian, 未知点 i 对应样本 j 的权重计算为

$$w_{ij}^{\tilde{\alpha}} = \sum_{S|i, j \in S} \left(-\frac{1}{|S|} \left(1 + (I_i - \mu_S) \left(\Sigma_S + \frac{\varepsilon}{|S|} E_3 \right)^{-1} (I_j - \mu_S) \right) \right), \quad (15)$$

其中 S 为包含点 i 与点 j 的特定集合. I_i 与 I_j 分别为点 i 与 j 的 3 维颜色向量, E_3 为 3×3 单位阵, Σ_S 为 S 的 3×3 协方差矩阵, μ_S 为 S 的 3×1 的均值向量, ε 为一个比较小的数. 这里根据 S 范围的大小, 将其分为以下 3 类: (a) 采用与 2.1 小节中的搜索方式类似的集合 S (不一定以 i 或 j 为起始搜索点), 这里采用 RGB 色彩空间, 距离分量的权重为一个级别 $w_i^s = 0.5$, 搜集 10 个最近邻样本 j , 且权重由式 (15) 取代了式 (2), 权重因子 (即与 $w_{ij}^{\tilde{\alpha}}$ 相乘的系数) 为 1; (b) 与 (a) 类似, 这里 $w_i^s = 0.1$, 最近邻样本数为 2, 权重因子为 0.1; (c) 半径为 1 的 3×3 大小的窗口 S , 此时与 closed form 方法相同, 权重因子为 7.

事实上, 式 (15) 体现在背景概率 $\tilde{\alpha}$ 而不是原始 α , 此时颜色线性模型仍然有效. 但该模型并不适用于前景, 因此公式 O_1 的后半部分 $w_{ij}^F = w_{ij}^{F^{\text{ind}}}$ 即采用与 2.2 小节相同的间接搜索前景的方式, 仅对前景各通道的前景做出进一步的平滑性处理, 且对各通道采用相同的权重. 这里距离分量的权重为 $w_i^s = 10$.

式 (13) 的第 2 部分 O_2 体现了 $\tilde{\alpha}$ 和 F 利用前 2 节的实体像素与混合像素的先验, 表述为

$$O_2(\tilde{\alpha}, F) = \sum_{i \in \Omega_{F'} \cup \Omega_{B'}} |\tilde{\alpha}_i - \tilde{\alpha}_i^{\text{pre}}| + \sum_{i \in \Omega_{U'}} \lambda f_i |\tilde{\alpha}_i - \tilde{\alpha}_i^{\text{mix}}| + \sum_{i \in \Omega_U} \sum_c |F_i^c - F_i^{\text{cpre}}|. \quad (16)$$

前 2 个部分为 $\tilde{\alpha}$ 对于实体像素和混合像素的先验, 其中 $\tilde{\alpha}^{\text{pre}}$ 与 $\tilde{\alpha}^{\text{mix}}$ 为 2.1 与 2.2 小节中求得的 α^{pre} 与 α^{mix} 的反向值, 第 3 个部分为 F 对于全体未知区域像素的先验, F^{pre} 为 2.2 小节中求得. 这里新的混合区域 $\Omega_{U'}$ 计算为

$$\Omega_{U'} = \Omega_U \cap \{i | \alpha_i^{\text{mix}} \in [0.2, 0.8]\}. \quad (17)$$

$\tilde{\alpha}$ 对于混合点的先验系数取 $\lambda = 0.05$, f_i 为 2.2 小节中通过颜色拟合误差求得的针对 α^{mix} 的信念值, 计算为

$$f = \exp(\|I - \alpha^{\text{mix}} F^{\text{pre}} - (1 - \alpha^{\text{mix}}) B^{\text{pre}}\|^2 / \sigma_2^2), \quad (18)$$

其中 $\sigma_2^2 = 0.1$. 信念值的实例如图 3(d) 所示.

式 (13) 的第 3 部分 O_3 体现了颜色拟合性平衡, 表述为

$$O_3(\tilde{\alpha}, F) = \sum_{i \in \Omega_U} \sum_c f_i |I_i^c - (B_i^{\text{cpre}} - F_i^{\text{cpre}}) \tilde{\alpha}_i - F_i^c|. \quad (19)$$

该项体现了求解分量 $\tilde{\alpha}$ 和 F 针对抠像公式 (1) 的拟合性, 采用反向值 $\tilde{\alpha}$ 是为了便于采用仿射方式求解, 这里若 F_i^{cpre} 与 F_i^c 相等, 则该式与原始抠像公式相同. 该部分的权重 $\theta = 10$.

在求解 $\tilde{\alpha}$ 后需要将其求反得到 α^* 值, 即 $\alpha^* = 1 - \tilde{\alpha}^*$. 结合 $\{F, \alpha^*\}$, 新背景 B' 及抠像公式 (1), 可以得到新合成图像 I' . 本小节的抠像实例以及合成实例如图 1 的第 3 层步骤及合成步骤所示.

3 实验

本文实验采用的是抠像评价系统^[25]中的 27 幅训练图像与 8 幅测试图像, 其中每幅图像包含了 3 种型号的 Trimap, 这里除了系统初始提供的“大型号”与“小型号”的 Trimap 之外, 还引入了一种“巨大型号”的 Trimap, 它是将大型号 Trimap 中的未知区域膨胀 10 个像素, 并保留面积小于 500 的前景和背景区域得到.

对于各种抠像方法, 在理论上需要比较 $\{\alpha, F\}$ 这 2 个分量. 其中 α 分量作为具有代表性地反映前景物体形状和透明度的最直接因素, 完全可以沿用传统 α 结果的比较方法. 然而 F 分量并不会随着前景物体边缘形状的变化而发生较大改变, 尤其对于 α 较小的像素, 对 F 分量进行比较通常会出现很大误差. 事实上, 抠像公式中的 αF 通常作为一个整体来表示 F 对 α 的依赖关系, 因此这里以 αF 的图像合成形式取代单独的 F 分量进行比较. 以下在各方法的 α 结果以及后续合成图像这 2 个层面分别进行比较.

本文并没有将结果提交至抠像评价系统^[25]中, 主要基于以下两点原因. 其一, 抠像系统中的很多传统仿射类方法均使用了针对单个像素逐一判断 (类似于采样方法) 的预处理步骤, 但本文主要是基于仿射类方法且并未采用基于采样方法的预处理步骤, 将本文方法与采用预处理步骤的方法得到的结果进行比较并不公平, 同时为了公平起见, 本文实验中将所有其他方法中的预处理步骤全部去掉, 包括 information flow 方法中的所有“Trimap 修剪”步骤, 以及 three-layer 方法中的“预处理”步骤; 其二, 本文主要面向合成应用, 并针对合成图像进行比较, 对 α 单独比较仅起到辅助作用.

3.1 α 结果的比较

由于本文的 α 是由仿射类方法求得, 因此公平起见这里也选取其他 4 种仿射类方法的 α 结果进行比较, 它们在后续合成图像的比较中也会用到, 包括 closed form (CLF), information flow (IFF), three-layer (THL), 以及本文第 1 步采用 KNN 搜索中的 HSV 特征计算初级 α 的分支方法 (HSV). 对于每种抠像方法, 针对评价系统中每种型号 Trimap 之下的 27 幅训练图像共对应 27 个 α 结果, 3 种型号 Trimap 共对应 81 个 α 结果. 于是对每种抠像方法, 其对应的某种 Trimap 型号 (或全体训练集) 的 MAD (平均绝对误差) 和 RMSE (均方根误差) 定义为

$$\begin{cases} \text{MAD}(\alpha) = \frac{1}{N} \sum_{\alpha} \sum_{i \in \Omega_U} |\alpha_i - \alpha_i^{\text{true}}|, \\ \text{RMSE}(\alpha) = \sqrt{\frac{1}{N} \sum_{\alpha} \sum_{i \in \Omega_U} (\alpha_i - \alpha_i^{\text{true}})^2}, \end{cases} \quad (20)$$

表 1 针对各训练图像及各型号的 Trimap, 5 种仿射类抠像方法得到的 α 结果与各真实值 α^{true} 之间的 MAD 与 RMSE 比较, 其中“巨大型”, “大型”, “小型”分别表示 3 种 Trimap 型号对应的各 27 幅图像的训练集, “全体”表示全体 81 幅图像的训练集

Table 1 MAD and RMSE comparisons between the α results obtained by five affinity-based matting methods and α^{true} values for each training image and each type of trimap, where “huge”, “large”, and “small” denote the three 27-image training subsets with respect to the 3 three Trimap sizes, and “entire” denotes the entire 81-image training dataset

Method	MAD ($\times 10^2$)				RMSE ($\times 10^2$)			
	Huge	Large	Small	Entire	Huge	Large	Small	Entire
CLF	9.39	7.88	6.89	8.37	18.39	16.43	14.37	16.98
HSV	5.35	6.09	7.35	6.02	14.41	14.93	15.68	14.85
IFF	6.16	5.87	5.81	5.99	13.90	12.71	11.61	13.06
THL	4.78	4.86	5.20	4.89	10.84	10.55	10.55	10.69
Ours	4.45	4.57	5.04	4.61	10.20	10.18	10.61	10.29

其中, N 为该 Trimap 型号 (或全体训练集) 对应的 27 幅图像 (或 81 幅图像) 中未知区域 Ω_U 的全体像素的个数, α 为当前方法针对 27 幅图像求得的 α 结果图像, α^{true} 为各图像的 α 真实值, i 为 α 图像中的像素索引. 上述 5 种方法的 MAD 和 RMSE 比较如表 1 所示. 显然, 本文方法的各 MAD 和 RMSE 基本均为最低. 同时, 若进一步加入预处理步骤, 在 α 评价系统^[25]中, 由于 information flow 与 three-layer 方法均排名较前, 因此也大致预测出本文的 α 结果会排在更加靠前的位置. 需要注意的是, 本文方法的 α 大致为 HSV 与 three-layer 方法的融合, 而且比二者结果都要好.

3.2 合成图像的比较

作为对比, 本文选取了除本文方法外的其他 8 种求解 $\{\alpha, F\}$ 方式, 共分为 2 大类. 第 1 类为已出现的 α 计算方法以及原文存在显式或隐式计算 $\{F, B\}$ 的方法: (1) 本文的中间步骤: 第 1 步求得的 α 与第 2 步中 $\{F, B\}$ 的计算方法 (HSV); (2) closed form 中的 α 及 $\{F, B\}$ 的计算方法 (简 CLF); (3) shared 中的 α 与 $\{F, B\}$ 的计算方法; (4) robust 中的 α 与 $\{F, B\}$ 的计算方法. 第 2 类为本文创造的、将之前提及的一些仿射类 α 计算方法与某些求解 $\{F, B\}$ 的方法相结合的方法: (5) information flow 中的 α 计算方法与 closed form 中计算 $\{F, B\}$ 方法相结合 (IFM+CLF); (6) three-layer 中的 α 计算方法与 Geodesic 中计算 $\{F, B\}$ 方法相结合 (THL+SAMP), 后者利用已求解的 α 以及采样类方式计算未知点 i 的 $\{F_i^*, B_i^*\}$:

$$\{F_i^*, B_i^*\} = \arg \min_{F_i, B_i} \|I_i - \alpha_i F_i - (1 - \alpha_i) B_i\|, \quad (21)$$

其中 F_i 与 B_i 为未知点 i 在已知前景与背景区域的边缘分别采集的 20 个最近邻样本; (7) information flow 中的 α 计算方法与 Geodesic 中计算 $\{F, B\}$ 方法相结合 (IFF+SAMP); (8) three-layer 中的 α 计算方法与 closed form 中计算 $\{F, B\}$ 方法相结合 (THL+CLF). 这里最后 2 种方法为第 5, 6 种方法将 α 与 $\{F, B\}$ 对结果进行了对调, 是为了体现相同 (或不同) α (或 F) 对最终合成结果的影响.

由于对 αF 的比较即为针对合成图像的比较, 因此需要建立合成图像数据集. 针对 27 幅训练图像、8 幅测试图像, 以及 3 种类型的 Trimap 型号, 实际上可建立的合成图像共有 $35 \times 3 = 105$ 幅. 本文根据绿色背景或自然图像背景建立了 2 组背景图像集, 每组各 105 幅图像. 在实验中, 每种抠像方法在各型号 Trimap 下得到的 $\{\alpha, F\}$ 结果与相对应数据集的背景 B' (绿背景或自然背景) 根据抠像公式 (1) 进行合成, 各合成数据集也均包含 105 幅合成图像. 同时, 每幅训练图像的真实 $\{\alpha, F\}$ 也与这些背景

图 4 (网络版彩图) αI 与 αF 合成方式的局部比较. (a)(d) 局部输入图像; (b)(e) αI 方式的合成图像; (c)(f) αF 方式的合成图像

Figure 4 (Color online) Local comparisons on compositing methods of αI and αF . (a)(d) Local images; (b)(e) synthetic images from αI ; (c)(f) synthetic images from αF

进行合成 (2 类数据集各包含 81 幅), 得到的真实合成图像作为比较基准, 其余 24 幅测试图像合成图 (无真实合成图) 则仅作为显示.

对于以上某个数据集上的某种特定 $\{F, B\}$ 计算方法, 在该方法下对合成图像 I' 的比较准则采用如下均方根误差 (RMSE) 的形式:

$$\text{RMSE}(I') = \sqrt{\frac{1}{N} \sum_{I'} \sum_i \sum_c (I'^c_i - I^{*c}_i)^2}, \quad (22)$$

其中, N 为训练集中针对某种型号的 Trimap (或全部 Trimap) 的全体像素的个数, I' 为当前方法在全体数据集或某种型号 Trimap 对应的子集中的某幅合成图像, i 为图像中的像素索引, c 为颜色通道索引, I^* 为合成图像 I' 对应的真实合成图像.

除了常规合成图像层面的比较外, 本文同时引入了 αI 层面的比较, 即用 αI 代替抠像公式 (1) 中的 αF 合成新图像, 合成公式为 $\alpha I + (1 - \alpha)B' = I'$, 这是一种比较传统且相对近似的方法^[14]. 前文提到, 很多抠像方法 (包括近期基于深度卷积网络的抠像类方法) 都没有计算 F 的方法, 在需要合成时, 该类方法通常采用原图 I 替代 F . 针对绿背景与一般背景, αI 合成方式形成的合成图像也各有 105 幅, 其中可用于数值比较的训练图各有 81 幅. 通常, 这种近似的方式在“硬边界” (即半透明像素变化快速的窄前景边界) 处会与真实效果大体一致, 但存在明显的问题在于半透明像素较多的 α 较小处容易将较多背景颜色带至合成图像中, 这也被称为“背景色溢出”现象. 采用 αI 的与 αF 合成 2 个例子如图 4 所示. 显然, 在前景边缘存在较多毛发处 (α 较小), αI 会将较多原图背景颜色 (此例为蓝色) 合成其中, 造成大量合成误差 (如图中毛发边缘部分), 而利用 αF 进行合成则不会出现此类问题.

显然, 由于原图像 I 与新背景 B' 均为已知, 各抠像方法在合成图 $\alpha I + (1 - \alpha)B'$ 中的变化量仅为 α , 因此采用 αI 合成方法的合成结果也可以模拟 α 结果之间的比较. 同时, 观察 αF 的合成方式相对 αI 的变化, 可以推测出 F 值相对 α 值对于合成效果的提升或降低. 这就是 αI 合成方式的意义.

表 2 与 3 分别显示了绿色背景与自然背景下采用 αI 与 αF 两种合成方法的 RMSE 比较, 体现于以下 3 点. (1) 总体表现. 显然, 在 3 种不同型号的 Trimap 对应的训练集以及全体训练集上, 本文在上述 2 种不同背景的合成方式下的 RMSE 均为最低, 表明本文在不同型号 Trimap 的输入之下均有最好的表现. 同时, 本文中间结果的效果仅次于本文最终结果, 说明初级 α^{pre} 结果与第 2 步求得的前景与背景结果已比较理想, 而最终优化步骤也起到了锦上添花的效果. (2) $\alpha F - \alpha I$ 值. 前文提到, αI 体现 α 单纯的合成能力, 而 αF 体现了二者的联合合成能力, αF 与 αI 之间的差异为负数体现了前景 F 对单纯 α 合成能力的进一步提升, 而正数表明了 F 对 α 合成能力的降低, 数值表明提升或降低的程度. 同时, 由于各方法的 F 值均经由 α 计算求得, 或者与 α 的计算存在很强的关联性, 因此 α 结果较好 (即 α 合成能力好) 的方法 F 结果同样很好, 表现在利用 F 后的合成对 α 合成的提升效果仍然较好 (负值), 例如 IFF+CLF (或 THL+CLF) 方法与 CLF 采用相同求解 F 的方法, 但由于 IFF+CLF (或 THL+CLF) 相比 CLF 对 α 计算更加准确, 因此较 CLF 在最终合成的性能上也有很大提高. (3) 在求

表 2 绿色背景下, 9 种 $\{\alpha, F\}$ 的计算方法生成的 2 种合成图像与真实合成图像的 RMSE 比较 (按 αF 方式的降序排列)

Table 2 RMSE comparisons (descending order by αF) between two kinds of synthetic images generated by seven $\{\alpha, F\}$ computation methods and true synthetic images under green backgrounds

Method	αI synthetic method				αF synthetic method				$\alpha F - \alpha I$
	Huge	Large	Small	Entire	Huge	Large	Small	Entire	
CLF	16.82	13.39	11.10	13.97	23.21	14.72	10.93	17.07	+3.10
Robust	19.85	14.56	11.55	15.70	21.88	15.18	11.12	16.66	+0.96
IFF+SAMP	13.95	11.57	10.10	11.98	16.77	11.96	9.39	13.07	+1.09
Shared	15.22	11.66	10.10	12.51	16.09	11.45	8.98	12.52	+0.01
THL+SAMP	11.87	10.44	9.71	10.71	15.32	10.96	8.89	12.02	+1.31
IFF+CLF	13.95	11.57	10.10	11.98	13.19	10.12	8.25	10.72	-1.26
THL+CLF	11.87	10.44	9.71	10.71	12.08	9.23	8.26	9.99	-0.72
Intermediate results	11.41	10.35	9.76	10.53	11.14	9.45	8.32	9.70	-0.83
Final results	11.02	10.21	9.71	10.33	10.02	8.81	8.21	9.05	-1.28
Truth values	8.11	8.11	8.11	8.11	0	0	0	0	-8.11

表 3 自然背景下, 7 种 $\{\alpha, F\}$ 的计算方法生成的 2 种合成图像与真实合成图像的 RMSE 比较 (按 αF 方式的降序排列)

Table 3 RMSE comparisons (descending order by αF) between two kinds of synthetic images generated by seven $\{\alpha, F\}$ computation methods and true synthetic images under natural backgrounds

Method	αI synthetic method				αF synthetic method				$\alpha F - \alpha I$
	Huge	Large	Small	Entire	Huge	Large	Small	Entire	
Robust	22.51	16.25	12.58	17.60	24.32	17.10	12.18	18.55	+0.95
CLF	18.78	15.15	11.82	15.51	24.60	16.69	11.55	18.41	+2.90
IFF+SAMP	15.56	12.46	10.67	13.05	18.07	12.65	9.95	13.97	+0.92
Shared	16.88	12.70	10.73	13.68	17.52	12.59	9.62	13.64	-0.04
THL+SAMP	12.98	11.31	10.27	11.57	16.07	11.64	9.47	12.69	+1.12
IFF+CLF	15.56	12.46	10.67	13.05	14.81	10.96	8.97	11.83	-1.22
THL+CLF	12.98	11.31	10.27	11.57	12.78	10.08	8.88	10.71	-0.86
Intermediate results	12.21	11.04	10.26	11.20	11.85	10.15	8.90	10.37	-0.83
Final results	11.92	10.94	10.20	11.04	10.95	9.62	8.78	9.82	-1.22
Truth values	8.11	8.11	8.11	8.11	0	0	0	0	-8.11

解 α 上, 仿射类方法中的 three-layer (THL) 比 information flow (IFF) 以及 closed form (CLF) 等方法要好; 在求解 F 上, 基于 CLF 的仿射类方法要比采样类方法要好, 例如 IFF+CLF (或 THL+CLF) 比 IFF+SAMP (或 THL+SAMP) 要好. 同时, $\{\alpha, F\}$ 全部采用采样类方法求解的 Robust 与 Shared 方法均效果不够理想. 本文方法充分结合了 THL 求解 α 的方法与 CLF 类求解 F 特性并加以进一步优化, 效果因此达到了最优.

图 5 显示了本文方法与其他 7 种方法 (不包括本文中间结果) 在绿色背景下合成图像的 3 个例子. 由于绿色背景没有任何纹理干扰, 因此较自然背景能够更清晰的体现图像细节. 前 2 个例子着重显示了一些混合像素的细节, 由于毛发等边缘处的混合像素对前景的变化非常平缓, 可以看作近似不

图 5 (网络版彩图) 训练集中绿色背景合成图像的实例

Figure 5 (Color online) Examples of synthetic images with green backgrounds in training dataset

变, 因此这 2 个例子主要以 α 变化为主, 即前景物体的轮廓或前景色彩的深与浅. 第 3 个例子着重显示了实体像素的细节, 除 α 外, 这里的前景实体也存在很多颜色的变化, 因此该例子中的前景 F 的色彩也非常重要. 除本文方法外, 各种方法大多会将一些原图像的背景纹理或颜色引入至合成图像中, 前景颜色也不如本文方法准确. 其中, Robust 方法中过于局限于局部采样的方式, 使得在未知区域过宽或前景存在空洞时效果不佳, Shared 方法由于采样范围较广且针对前景加入了平滑措施, 因此效果较 Robust 好一些. CLF 方法采用了局部性仿射类方法, 同样无法准确计算透明度与前景颜色. IFF 求解 α 的方法中, 无论采用仿射类 IFF+CLF 还是采样类 IFF+SAMP 求解 F , 由于在透明度的搜索范围较本文方法存在差距, 同时对前景的计算仍然采用 CLF 方法, 因此合成效果仍然不够好. 在 THL 求

图 6 (网络版彩图) 训练集中自然背景合成图像的实例

Figure 6 (Color online) Examples of synthetic images with natural backgrounds in training dataset

解 α 的方法中, 采用采样方法求解 F 的 THL+SAMP 并不理想, 而仿射类方法求解的 THL+CLF 与本文方法相似, 因此结果与本文方法较为相近. 本文方法在各实例的前景细节部分的颜色值、分布情况和清晰度均与真实值最为接近.

图 6 显示了本文方法与其他 7 种方法在自然背景下合成图像的 2 个例子. 第 1 个例子的第 1 个细节为前景左上部分的实体像素, 包括不透明的白纸与铅笔, 第 2 个细节为前景右侧半透明的网格, 第 2 个例子的细节为前景布条的末端. 针对上述局部例子, 本文方法较其他方法均能够更加干净与完整的体现出各个局部细节. 在整体上, 第 1 个例子的前景网状杯子颜色变化不大并以整体 α 变化为主, 而第 2 个例子中实体布条的整体色彩以前景颜色 F 为主, 本文方法在对于这 2 个例子较其他方法均能够真实反映出它们各自的特点.

图 7 显示了 2 个完全半透明前景物体的例子, 由于均属于测试集, 因此没有真实值进行比较. 此时由于前景颜色非常平缓, 且采用 CLF 方式求解 F 的方法非常连续, 不会产生锯齿和断层的现象, 因此单纯的视觉感受上, 本文以及采用后续 CLF 方式所呈现合成图像的视觉体验最好.

4 结论

数字图像抠像问题通常与合成问题相关联, 二者组成了不可分割的整体, 脱离了合成问题的单独计算透明度的抠像方法属于单纯理论层面的研究, 无法进一步拓展至实际应用层面中. 同时, 一些传统计算前景颜色方法的实际合成效果也并不理想. 本文将抠像问题视为合成问题的子问题, 采用仿射类方式同时计算输入图像对应的透明度值和前景颜色值, 将抠像问题引至实际合成应用领域中, 提升了抠像问题的实际应用价值.

图 7 (网络版彩图) 测试集中半透明图像对应的合成图像实例

Figure 7 (Color online) Examples of synthetic images from semi-transparent objects in testing dataset

针对实体像素与混合像素这两类抠像问题中的经典像素集合, 本文将求解问题分为 3 个步骤. 前 2 个步骤分别计算实体像素及混合像素的先验并应用于后续步骤, 最终步骤同时求解出透明度和前景颜色值. 除先验外, 本文充分利用了仿射类方法中像素相关性的特点, 在各步骤中对透明度和前景颜色同时建立相关仿射矩阵, 作为先验数据项的补充. 此外, 本文进一步利用抠像公式对求解整体进行辅助性拟合平衡. 在实验中, 在传统的透明度比较之外, 本文建立了绿色背景及自然背景下的一般性合成数据集, 以及专门针对透明度的合成数据集, 在已出现的合成类方法及本文设计的一些新方法的基础上, 将本文方法与这些方法在上述合成数据集上分别进行比较. 实验表明, 本文方法不仅对于传统透明度的计算中具有优势, 而且对于合成图像的质量同样优于其他方法.

参考文献

- 1 Xu N, Price B, Cohen S, et al. Deep image matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 2970–2979
- 2 Zhang Y K, Gong L X, Fan L B, et al. A late fusion CNN for digital matting. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 2019. 7469–7478
- 3 Chen Q, Ge T Z, Xu Y Y, et al. Semantic human matting. In: Proceedings of the 26th ACM International Conference on Multimedia, Seoul, 2018
- 4 Yao G L, Zhao Z J, Liu S H. A comprehensive survey on sampling-based image matting. Comput Graph Forum, 2017, 36: 613–628
- 5 Wang J, Cohen M F. Optimized color sampling for robust matting. In: Proceedings of IEEE Conference on Computer

- Vision and Pattern Recognition, 2007. 1–8
- 6 Shahrian E, Rajan D, Price B, et al. Improving image matting using comprehensive sampling sets. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 636–643
 - 7 Varnousfaderani E S, Rajan D. Weighted color and texture sample selection for image matting. IEEE Trans Image Process, 2013, 22: 4260–4270
 - 8 Karacan L, Erdem A, Erdem E. Alpha matting with KL-divergence-based sparse sampling. IEEE Trans Image Process, 2017, 26: 4523–4536
 - 9 Gastal E S L, Oliveira M M. Shared sampling for real-time alpha matting. Comput Graph Forum, 2010, 29: 575–584
 - 10 Chuang Y Y, Curless B, Salesin D H, et al. A Bayesian approach to digital matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, 2001. 264–271
 - 11 Bai X, Sapiro G. Geodesic matting: a framework for fast interactive image and video segmentation and matting. Int J Comput Vis, 2009, 82: 113–132
 - 12 He K M, Rhemann C, Rother C, et al. A global sampling method for alpha matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado, 2011. 2049–2056
 - 13 Johnson J, Varnousfaderani E S, Cholakal H, et al. Sparse coding for alpha matting. IEEE Trans Image Process, 2016, 25: 3032–3043
 - 14 Li X Q, Li J D, Lu H. A survey on natural image matting with closed-form solutions. IEEE Access, 2019, 7: 136658
 - 15 Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting. IEEE Trans Pattern Anal Mach Intell, 2008, 30: 228–242
 - 16 Lee P, Wu Y. Nonlocal matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, 2011. 2193–2200
 - 17 Chen Q F, Li D Z Y, Tang C K. KNN matting. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 2175–2188
 - 18 Chen X W, Zou D Q, Zhou S Z Y, et al. Image matting with local and nonlocal smooth priors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 1902–1907
 - 19 Shi Y F, Au O C, Pang J H, et al. Color clustering matting. In: Proceedings of IEEE International Conference on Multimedia and Expo, San Jose, 2013
 - 20 Aksoy Y, Aydin T O, Pollefeys M. Designing effective inter-pixel information flow for natural image matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017
 - 21 Li C, Wang P, Zhu X Y, et al. Three-layer graph framework with the sumD feature for alpha matting. Comput Vis Image Und, 2017, 162: 34–45
 - 22 Grady L, Schiwietz T, Aharon S, et al. Random walks for interactive alpha matting. In: Proceedings of International Conference on Visualization, Imaging and Image Processing, Benidorm, 2005. 423–429
 - 23 Yao G L. A survey on pre-processing in image matting. J Comput Sci Technol, 2017, 32: 122–138
 - 24 Feng F J, Huang H, Wu Q X, et al. An alpha matting algorithm based on collaborative swarm optimization for high-resolution images. Sci Sin Inform, 2020, 50: 424–437 [冯夫健, 黄翰, 吴秋霞, 等. 基于群体协同优化的高清图像前景遮罩提取算法. 中国科学: 信息科学, 2020, 50: 424–437]
 - 25 Rhemann C, Rother C, Wang J, et al. A perceptually motivated online benchmark for image matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009. 1826–1833

A multi-level affinity-based matting and compositing method based on a joint solution of foreground and transparency

Guilin YAO^{1,2}, Biying ZHANG^{1,2}, Yanrong ZHANG^{1,2} & Xiaodong SU^{1,2*}

1. School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China;

2. Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin 150028, China

* Corresponding author. E-mail: suxd@hrbcu.edu.cn

Abstract Digital image matting and compositing is a classical problem in virtual reality. According to the matting formula, the matting problem needs to accurately extract the alpha value and the color value of the foreground object from the input image, while the compositing problem only uses the above matting results to combine with the new background image to make it a new synthetic image. However, at present, plenty of matting algorithms are only limited to the calculation of alpha value in matting, and ignore the calculation of foreground color used in the subsequent compositing steps. In fact, the above way is only limited to the theoretical research, and is separated from the compositing application level in the matting problem. Meanwhile, the effect of other algorithms for simultaneous foreground color calculation is still unsatisfying, and there is still a lack of compositing image data sets and comparison of compositing images. In this paper, we make full use of the advantages of the flexible search range in the affinity-based method, and use the three-level structure of the opaque pixels and the mixed pixels for hierarchical processing to jointly solve the final alpha value and foreground color. Among them, each level uses the forward level priors along with the pixel correlation of the affinity method and color fitting constraints, and finally solves the problem in a closed form solution. In the experiments, this paper not only compares the results in alpha level, but also introduces the comparison of compositing image level in the field of digital matting for the first time. In the compositing layer, white backgrounds and corresponding natural backgrounds are both introduced for the input image and trimap of each size. At the same time, in each method, the real alpha value and foreground color are generally composited with these backgrounds, and a simple alpha level composition is also carried out. The experimental results show that, in terms of the quantitative evaluation and visual quality, the final alpha results and the above compositing effects of this method are better than the state-of-the-art matting and compositing methods as well as some newly generated methods.

Keywords image matting, image compositing, affinity-based matting, closed form solution, opaque and mixed pixels

Guilin YAO was born in 1980. He received his B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology. Currently, he is an associate professor at the School of Computer and Information Engineering, Harbin University of Commerce. His research interests are image processing, artificial intelligence, and computer vision.

Biying ZHANG was born in 1976. He received his M.E. degree in computer software and theory from Harbin Engineering University, and Ph.D. degree in computer architecture from Harbin Institute of Technology. Currently, he is an associate professor at Harbin University of Commerce. His main research interests include distributed computing, machine learning, and big data processing.

Yanrong ZHANG was born in 1982. She received her Ph.D. degree from Northeast Forestry University, Harbin. Currently, she is an associate professor at the School of Computer and Information Engineering, Harbin University of Commerce. Her main research interests are E-commerce and artificial intelligence.

Xiaodong SU was born in 1965. He is a professor at Harbin University of Commerce and is the member of council of the Heilongjiang Computer Society, vice president of Electronics Experiment Teaching Research Council of Universities in Heilongjiang, and the managing director of Heilongjiang Software and Information Services Association. His current research interests include signal and information processing and computer vision.