

DOI: 10.11992/tis.201809014

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20181227.1144.004.html>

局部自适应输入控制的随机游走抠图

陈秋凤^{1,2}, 申群太²

(1. 福建农林大学 计算机与信息学院, 福建 福州 350002; 2. 中南大学 信息科学与工程学院, 湖南 长沙 410083)

摘 要: 针对传统性抠图算法中, 非完全正确用户标注及不精确超像素分割造成的信息误扩散, 以随机游走算法为基础, 提出带软性约束的抠图算法。通过对扩展 Dirichlet 问题的推导, 指出带软约束的随机游走与部分自吸收随机游走概率的关联性。以吸收概率为指导, 在传统相似扩散所构建的图模型上, 根据局部窗口内特征矩阵的秩与方差设计了输入控制矩阵, 使得信息扩散的过程能够跟随图像的局部特征进行自适应扩散。最后将软约束随机游走应用到单帧双层抠图及视频抠图中。实验表明, 所提算法具有信息远距传播能力和良好的容错性能, 尤其在用户标注不够充分的情况下能够取得更加优良的抠图结果。

关键词: 抠图; 视频抠图; 随机游走; 软性约束; 吸收概率; 局部模型; 输入控制; 自适应

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2019)05-1007-10

中文引用格式: 陈秋凤, 申群太. 局部自适应输入控制的随机游走抠图 [J]. 智能系统学报, 2019, 14(5): 1007-1016.

英文引用格式: CHEN Qiufeng, SHEN Quntai. Random-walk matting with local adaptive input control[J]. CAAI transactions on intelligent systems, 2019, 14(5): 1007-1016.

Random-walk matting with local adaptive input control

CHEN Qiufeng^{1,2}, SHEN Quntai²

(1. College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: In traditional image-matting algorithms, incomplete user labeling and inaccurate super-pixel segmentation lead to the incorrect propagation of information. To solve this problem, we propose the use of soft constrained matting based on a random-walk algorithm. Through the derivation of the extended Dirichlet problem, we identify the relationship between a random walk with soft constraint and the probability of a partial self-absorption random walk. Guided by the absorption probability, an input control matrix is designed according to the rank and variance of the feature matrix in the local window. This is performed via a graph model that was constructed using traditional similarity diffusion such that the process of information diffusion could follow the local image features to realize adaptive diffusion. Finally, we applied the soft constrained random walk to single-frame two-layer image and video matting. The experimental results reveal that the proposed algorithm can transmit information over long distances and has good fault tolerance. In addition, it can achieve better image matting results, particularly in cases wherein user labeling is insufficient.

Keywords: matting; video matting; random walk; soft constrained; absorption probability; local model; input control; adaptive control

抠图是按照不透明度将感兴趣物体, 从图像或视频序列中精确分离出来的一种图像处理技

术^[1-2]。抠图是一个高病态问题, 需要用户提供一定的标注信息进行求解。目前的单帧抠图算法都要求用户输入的标注信息完全正确, 并采用大数值的输入控制参数, 以迫使输出值与用户标注值严格相同^[3-12]。然而过强的输入约束, 使得信息传

收稿日期: 2018-09-11. 网络出版日期: 2018-12-27.

基金项目: 国家自然科学基金项目 (61473318, 60974048).

通信作者: 陈秋凤. E-mail: chenqiufeng0204@126.com.

播只与标注区域的边界相关,传播距离有限;在多层抠图算法中^[13],通过超像素来构建不同层级的图像,虽能够提高算法运算速度,但受超像素预分割精度的影响,高层计算出的结果在向下层传递标注时也会造成错误的初始值。而视频抠图是单帧抠图在图像序列流上的扩展^[14-18],帧间标注信息的传递尤为重要。目前视频抠图多数采用半自动标注的方式,通过帧间传播策略将关键帧上的标注信息依次向后续帧传递。虽有学者^[15,18]利用光流信息提高了帧间连续性,但依然采用的是硬约束的方式,因此要求传播产生的三分图对前景边界有良好的包围性并严格正确,这使得后续帧的三分图产生过程复杂,影响了算法的可扩充性和快速性。针对三分图标注的产生方法,不少学者也作了进一步的研究,但仍然是建立在初始标注完全正确的基础上^[19-20]。

综上所述,传统算法采用的硬约束方式使得抠图效果严重依赖于所采用标记的准确性,对用户输入要求高。为此,本文在随机游走算法基础上,提出了软约束随机游走算法 (soft-constrained random walk, SCRW),使得输入控制矩阵能够根据图像颜色分布特性进行自适应调整。

1 随机游走分割算法

随机游走是一种具有马尔科夫链性质的特殊布朗运动,在给定的图和一个出发点上,信息以一定的概率随机地移动到邻居节点上。借助电势理论,Grady^[21]指出图像分割过程实际上是求解带边界条件的 Dirichlet 问题。首先建立一张自然图像对应的无向图模型 $G=(V,E)$, 节点 V 表示图像中的像素点, E 表示连接两个节点的边, $W=[w_{ij}]_{n \times n}$ 为相似度矩阵, w_{ij} 表示节点 i 与节点 j 的相似度,定义节点的度矩阵 D , 其对角元素为 $d_i = \sum_j w_{ij}$ 。给定有 n 个像素的数据集 $\chi = \{x_1, x_2, \dots, x_n\}$, 图像分割的目标是将数据集 χ 分成 k 类。设 χ_M 为用户预先指定的种子点集 (每一类至少有一个已标注的种子点), χ_U 为未标注点集, 则原数据集可表示为 $\chi = [\chi_M^T \chi_U^T]^T$ 。从每个未标注点出发, 分别计算该未知点到 k 类标注点的首达概率, 并根据最大概率将该点划分到相应的类别, 从而实现图像的分割。记图的拉普拉斯矩阵为 $L=D-W$, 则

$$L_{ij} = \begin{cases} \sum_j w_{ij}, & i=j \\ -w_{ij}, & i \neq j, j \in \Omega_i \\ 0, & \text{其他} \end{cases}$$

式中 Ω_i 表示像素点 i 的空间近邻集合。令

$\mathbf{p}^k = [p_1^k \ p_2^k \ \dots \ p_n^k]^T$ 表示数据集对第 k 个类别的首达概率, 并将 \mathbf{p}^k 改写为 $\mathbf{p}^k = [(\mathbf{p}_M^k)^T \ (\mathbf{p}_U^k)^T]^T$, \mathbf{p}_M^k 为已标注的种子点概率, 取值为

$$p_i^k = \begin{cases} 1, & x_i \text{ 属于第 } k \text{ 类} \\ 0, & \text{其他} \end{cases}$$

由文献 [21] 知, 随机游走分割算法的 Dirichlet 问题也可写成两顶点概率差值的加权和形式:

$$D[\mathbf{p}^k] = \frac{1}{2} \sum_{e_{ij} \in E} w_{ij} (p_i^k - p_j^k)^2 \quad (1)$$

求解可得未标注点的首达概率 \mathbf{p}_U^k , 即

$$\mathbf{p}_U^k = -\mathbf{L}_U^{-1} \mathbf{B}^T \mathbf{p}_M^k \quad (2)$$

2 带输入控制的随机游走抠图算法

2.1 目标函数规则化约束与输入控制

在传统双层抠图中, 其求解的目标函数不但要求两邻近像素点间的 α 值最大程度地符合建立的图模型, 保持局部相似性, 也要求输出 α 值与原始给定值相一致^[3-12]。因而其目标函数通常包含有平滑项和数据项两个部分:

$$J = \min \frac{1}{2} \left(\sum_{i=1}^n \sum_{j \in \Omega_i} w_{ij} (\alpha_i - \alpha_j)^2 + \lambda \sum_{i \in S} (\alpha_i - \tilde{\alpha}_i)^2 \right) \quad (3)$$

式中: α_i 为点 i 的不透明度; $\tilde{\alpha}_i$ 为待求不透明度的输入初始值; S 为约束集合; λ 为输入控制参数, 其值越大表示所求 α 值与原输入的一致性越高。式 (3) 等号左边第 2 项可看作针对输入信息的规则化项, λ 为规则化因子。

结合原始的随机游走算法, 将传统双层抠图算法扩展为多图层抠图, 像素点到 k 类种子的首达概率定义为第 k 个图层的不透明度, 并与式 (3) 的 λ 相区别, 此处为每个点取不同的输入控制参数 h_i , 将抠图目标函数转化成带规则化输入信息约束的扩展 Dirichlet 问题, 则

$$D[\alpha^k] = \frac{1}{2} \left(\sum_{i=1}^n \sum_{j \in \Omega_i} w_{ij} (\alpha_i^k - \alpha_j^k)^2 + \sum_{i \in S} h_i (\alpha_i^k - \tilde{\alpha}_i^k)^2 \right) \quad (4)$$

式中: α_i^k 、 α_j^k 表示点 i 、 j 在第 k 个图层的不透明度值; $\tilde{\alpha}_i^k$ 为用户输入的信息或初始值; h_i 为第 i 个像素点的输入控制参数, 其数值的大小与信息的传播距离相关。当 $k > 2$ 时, 式 (4) 为多层抠图; 当 $k = 2$ 时, 式 (4) 为单纯的前景提取。若将式 (1) 中的首达概率 \mathbf{p}^k 定义为不透明度 α^k , 再与式 (4) 进行对比, 不难发现式 (1) 的传统随机游走算法少了规则化约束项。实际上随机游走分割是将种子点看作理想电源, 内阻为无限小, 即假设 λ 为无穷大, 从而才能对拉普拉斯矩阵进行拆分, 将数据集分成已标注和未标注两个部分求解。而式 (3) 的传统抠图算法的目标函数也都是将输入控

制参数 λ 取为大值^[3-12], 并且每个点的输入控制参数值相等。若忽略各算法所构建图模型的差异性 (w_{ij} 看成相同), 当各抠图算法的输入控制参数 $\lambda \rightarrow \infty$ 时, 很容易得知各抠图算法与随机游走分割算法具有等价性。

λ 值是否越大越好? 其实不然。因 λ 值取为大值存在以下问题:

1) 标注信息从已知区域向未知区域扩散时, 未知区域只能接受已标注区域的边界信息, 扩散过程只依赖于局部相似关系建立的图模型, 而与已知区域内部的其他信息无关。尤其是标注信息不足时, 局部模型的小窗口特性限制了输入信息的传播距离。

2) 为了使目标函数值最小, 传统算法会迫使 α 的取值在已知区域和未知区域之间平滑过渡。当输入的信息不完全正确时, 过大的 λ 值也会将错误传递给其他像素, 且无法进行自修正。

针对 λ 取大值存在的问题, 以下对输入控制参数 h_i 取小值时的特性进行分析。

2.2 输入控制参数 h_i 取小值时转移概率分析

将式 (4) 写成矩阵形式, 则有

$$D[\alpha^k] = \frac{1}{2} \left((\alpha^k)^T L(\alpha^k) + (\alpha^k - \tilde{\alpha}^k)^T H(\alpha^k - \tilde{\alpha}^k) \right) = \frac{1}{2} \left(\left[(\alpha_l^k)^T (\alpha_u^k)^T \right] \begin{bmatrix} L_l & B \\ B^T & L_u \end{bmatrix} \begin{bmatrix} \alpha_l^k \\ \alpha_u^k \end{bmatrix} + \dots + \left(\begin{bmatrix} \alpha_l^k \\ \alpha_u^k \end{bmatrix} - \begin{bmatrix} \tilde{\alpha}_l^k \\ \tilde{\alpha}_u^k \end{bmatrix} \right)^T \begin{bmatrix} H_l & 0 \\ 0^T & H_u \end{bmatrix} \left(\begin{bmatrix} \alpha_l^k \\ \alpha_u^k \end{bmatrix} - \begin{bmatrix} \tilde{\alpha}_l^k \\ \tilde{\alpha}_u^k \end{bmatrix} \right) \right) \quad (5)$$

式中: 对角矩阵 $H = \text{diag}\{h_1, h_2, \dots, h_n\}$; α^k 表示第 k 个图层的不透明度值; $\tilde{\alpha}^k$ 为给定的初值, 并将其分为已知和未知两部分用下标 u, l 表示。矩阵 H 对角元素 h_i 的大小表明了对原始输入值的遵从程度。为了达到更远的传播距离和避免不准确输入信息对输出结果的不良影响, 本文输入控制参数的取值较小, 相比于传统抠图算法硬约束下的 λ 值取大值, 式 (5) 可以看成一种带软约束的随机游走算法。将式 (5) 求微分得到优化的输出结果为

$$\alpha_u^k = (L_u + H_u)^{-1} (-B^T \alpha_l^k + H_u \tilde{\alpha}_u^k) = (L_u + H_u)^{-1} (-B^T \alpha_l^k) + (L_u + H_u)^{-1} H_u \tilde{\alpha}_u^k \quad (6)$$

当 H_u 取小数值时, $(L_u + H_u)^{-1} \approx L_u^{-1}$, 式 (6) 第一项与式 (2) 的无约束随机游走一致; 第二项与输入初始信息 $\tilde{\alpha}_u^k$ 相关, 令变换矩阵 $A = (L_u + H_u)^{-1} H_u$, 则有

$$A = (L_u + H_u)^{-1} H_u = (L_u + H_u)^{-1} (D_u + H_u) (D_u + H_u)^{-1} H_u = ((D_u + H_u)^{-1} (D_u - W_u + H_u))^{-1} (D_u + H_u)^{-1} H_u = (I - (D_u + H_u)^{-1} W_u)^{-1} (D_u + H_u)^{-1} H_u \quad (7)$$

式 (7) 等价于 $A - (D_u + H_u)^{-1} W_u A = (D_u + H_u)^{-1} H_u$,

将该式拆开, 并写成单个元素的形式为

$$a_{ii} = \frac{h_i}{h_i + d_i} \times 1 + \sum_{j \neq i} \frac{w_{ij}}{h_i + d_i} a_{ij} \quad (8)$$

$$a_{ij} = \sum_{j \neq k} \frac{w_{kj}}{h_i + d_i} a_{jk}, \quad i \neq j \quad (9)$$

由文献 [22] 可知, 式 (8)、式 (9) 正是部分吸收随机游走算法 (partial absorption random walk, PARW) 的基本形式, 故 A 为吸收概率矩阵, a_{ii} 表示信息在节点 i 的自吸收概率, a_{ij} 表示节点 i 的信息被近邻节点 j 吸收的概率。根据式 (8)、式 (9) 的分析, 可得到结论 1 和结论 2。

结论 1 由于吸收概率矩阵 A 中的元素在 0、1 之间, 且各行加和为 1, 因此输入信息 $\tilde{\alpha}_u^k$ 在图模型上的扩散有稳定解, $\tilde{\alpha}_u^k$ 所提供的信息将被图完全吸收。

结论 2 自吸收概率 a_{ii} 的大小与输入信息的局部相似扩散距离有关, 自吸收概率 a_{ii} 越小, 传播距离越远, 反之距离越近。

由于本文所提带软性约束条件的随机游走 SCRW 与部分吸收式随机游走算法 PARW 本质上是相同的, 此时图中节点 i 到节点 j 的转移概率为

$$p_{ij} = \begin{cases} \frac{h_i}{h_i + d_i}, & i = j \\ \frac{w_{ij}}{h_i + d_i}, & i \neq j \end{cases} \quad (10)$$

式 (10) 表明, 初始信息有 $h_i/(h_i + d_i)$ 的概率停留在原节点上, 而由 $w_{ij}/(h_i + d_i)$ 概率转移到相邻节点 j , 转移到相邻节点的总概率为 $d_i/(h_i + d_i)$ 。

3 输入控制矩阵设计

3.1 信息流扩散与图像局部模型

由上可知, 节点的自吸收概率与 d_i, h_i 相关, 在不同类型的图 $G = (V, E)$ 中, 信息流的扩散也不相同。以下将图 $G = (V, E)$ 分为非归一化图模型和归一化图模型两种进行讨论。

在非归一化图模型中, 若各点的输入控制参数取为相同的值, 即 $H = \gamma I$ (γ 为常量) 时, 图像点 i 的自转移概率为 $p_{ii} = \gamma/(\gamma + d_i)$, p_{ii} 随 d_i 单调递减。由于在图像边界内部像素点间相似度高, 节点在边界内的 d_i 值比边界处大, 导致 p_{ii} 在边界内的吸收概率低于边界, 当信息流到边界处时将会被节点高概率地吸收, 从而防止标注信息的扩散超过边界。非归一化矩阵 $H = \gamma I$ 之所以能保持信息大部分在边界内被吸收, 主要得益于图结构上各节点度 d_i 的差异性。

而在归一化图模型中, 各点的度都为 d_0 , 信息

流的自转移概率为 $h_i/(h_i + d_0)$, 若各点的输入控制参数也取为相同的值, 则各节点间的自转移概率都相同。在传统的 CF 抠图^[10]算法中, 基于像素的节点连接度为 $w_{ij} \propto (I_i - \mu_k)(I_j - \mu_k)/(\sigma_k^2 + \varepsilon)$, 若窗口半径为 r , 经过运算可知各节点的度相等, 即 $d_i = d_0 = (2r + 1)^2$ 。从 w_{ij} 的计算公式可以看出, CF 算法相当于对原数据进行了 0 均值的归一化操作, 归一化丢失了图像各窗口间颜色变化的差异性, 因此若将输入控制矩阵 \mathbf{H} 取为 $\gamma \mathbf{I}$, 则自转移概率 p_{ii} 为常数, 边界内和边界外的节点对信息流具有相同的吸收概率, 信息流由于没有边界吸收特性, 将出现误扩散。

因此, 在节点度 d_i 相同的归一化图模型中, 设计具有边界吸收特性的转移概率, 则要求各节点的输入控制参数 h_i 能够根据图像的局部分布特征而变化。以下探讨局部窗口内前背景颜色模型的分布特征, 及相应模型下 α 值相似变换的自由度; 并以此为基础来确定 h_i 值。

根据图像特性^[23], 在围绕图像点 i 的窗口中, 可将前景颜色和背景颜色的关系分为 4 种模型:

1) 线-线模型: 前景和背景都是线性变化, 则存在颜色常量 $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{B}_1, \mathbf{B}_2)$, 使窗口中点 j 的前景和背景为 $\mathbf{F}_j = \beta_j^F \mathbf{F}_1 + (1 - \beta_j^F) \mathbf{F}_2$, $\mathbf{B}_j = \beta_j^B \mathbf{B}_1 + (1 - \beta_j^B) \mathbf{B}_2$, 将 \mathbf{F}_j 、 \mathbf{B}_j 带入抠图前背景耦合公式 $\mathbf{I}_j = \alpha_j \mathbf{F}_j + (1 - \alpha_j) \mathbf{B}_j$, 则 $\mathbf{I}_j = \alpha_j [\beta_j^F \mathbf{F}_1 + (1 - \beta_j^F) \mathbf{F}_2] + (1 - \alpha_j) [\beta_j^B \mathbf{B}_1 + (1 - \beta_j^B) \mathbf{B}_2]$ 。

由于图像 RGB3 通道分别符合线性关系, 记 3×3 矩阵 \mathbf{Q} 的对应取值为 $[\mathbf{F}_2 + \mathbf{B}_2 \quad \mathbf{F}_1 - \mathbf{F}_2 \quad \mathbf{B}_1 - \mathbf{B}_2]$, 则 \mathbf{I}_i 与 \mathbf{Q} 的关系改写为

$$\mathbf{Q} \begin{bmatrix} \alpha_j \\ \alpha_j \beta_j^F \\ (1 - \alpha_j) \beta_j^B \end{bmatrix} = \mathbf{I}_j - \mathbf{B}_2$$

式中 $\mathbf{I}_j = [\mathbf{I}_j^F \quad \mathbf{I}_j^B \quad \mathbf{I}_j^T]^T$, 取 $\alpha_i^F \alpha_i^B \alpha_i^T$ 为 \mathbf{Q}^{-1} 的第 1 行, 则存在 4 个自由度的相似变换:

$$\alpha_j = \alpha_i^F \mathbf{I}_j^F + \alpha_i^B \mathbf{I}_j^B + \alpha_i^T \mathbf{I}_j^T + b_i$$

2) 点-线模型: 前景或是背景其中之一退化为点模型, 在窗口内取值为常数, 不妨设前景为常量, 背景呈线性, 则 $\mathbf{F}_j = \mathbf{F}$, $\mathbf{B}_j = \beta_j^B \mathbf{B}_1 + (1 - \beta_j^B) \mathbf{B}_2$, 代入线性组合公式得

$$\mathbf{I}_j = \alpha_j \mathbf{F} + (1 - \alpha_j) (\beta_j^B \mathbf{B}_1 + (1 - \beta_j^B) \mathbf{B}_2) = \alpha_j (\mathbf{F} - \mathbf{B}_2) + (1 - \alpha_j) \beta_j^B (\mathbf{B}_1 - \mathbf{B}_2) + \mathbf{B}_2$$

推导可知存在自由度为 3 的变换:

$$\alpha_j = \alpha_i^F \mathbf{I}_j^F + \alpha_i^B \mathbf{I}_j^B + \alpha_i^T \mathbf{I}_j^T$$

3) 点-点模型: 前景和背景都取值为常数的点模型, 即 $\mathbf{F}_j = \mathbf{F}$, $\mathbf{B}_j = \mathbf{B}$, 此时 $\mathbf{I}_j = \alpha_j \mathbf{F} + (1 - \alpha_j) \mathbf{B} = (\mathbf{F} - \mathbf{B}) \alpha_j + \mathbf{B}$, 则自由度为 2 的变换为

$$\alpha_j = \alpha_i^F \mathbf{I}_j^F + \alpha_i^B \mathbf{I}_j^B$$

4) 单点模型: 前景和背景取为相同常数, 即 $\mathbf{I}_j = \mathbf{F} = \mathbf{B}$, α_j 为任意常量, 这种特殊情况下, 自由度为 1, 窗口内各点颜色值相同, 一般是在图像连续平滑区域。

3.2 局部自适应输入控制矩阵设计

记局部窗口内的特征矩阵为 $\mathbf{G}_i = [\mathbf{I}_j^F \quad \mathbf{I}_j^B \quad \mathbf{I}_j^T \quad 1]_{j \in \Omega_i}$, 当前背景颜色取线-线、点-线、点-点及单点模型时, 由 3.1 节知颜色特征到 α 值变换的自由度为 4、3、2、1, 因而相应模型下特征矩阵的秩也为 4、3、2、1。矩阵的秩越低, 表示特征向量间的相关性越大, 窗口内像素点间具有高的相似度, 这使得隐含的未归一化前的节点度值 d_i 更大, 若图模型未进行归一化, 则不同的 d_i 能够使信息流根据图像特征进行扩散, 因而输入控制参数可取为相同的值; 若图模型进行了归一化, 当窗口内的方差较小时, 局部区域颜色变化平滑, 则希望有较小的 h_i 来使得当前节点的自吸收概率低, 从而信息流有更大的流动性, 反之则 h_i 取较大的值。总之, 参数 h_i 设计的目标是使信息流根据图像的局部特征得到合适的扩散, 即在模型简单颜色平滑区域的扩散程度高, 复杂区域的扩散程度低。综上所述, 本文采用的 h_i 计算公式为

$$h_i = \begin{cases} \gamma_1, & \mathbf{G} = (\mathbf{V}, \mathbf{E}) \text{ 是非归一化图} \\ \gamma_2 \times \text{rank}(\mathbf{G}_i) \times \frac{\sigma_i}{\bar{\sigma}}, & \mathbf{G} = (\mathbf{V}, \mathbf{E}) \text{ 是归一化图} \end{cases} \quad (11)$$

式中: γ_1 、 γ_2 为全局输入控制常数, 为避免过强输入限制作用, 一般将其取为小值; $\text{rank}(\mathbf{G}_i)$ 为特征矩阵的秩; σ_i 为窗口内颜色方差, 整张图像的平均方差为 $\bar{\sigma} = \sqrt{\sum_i (\sigma_i)^2 / n}$ 。当建立的是非归一的图模型时, 各节点的输入控制参数 h_i 取相同的值, $p_{ii} = \gamma_1 / (\gamma_1 + d_i)$, 各节点度 d_i 的差异性使得在未归一化图上的扩散自然具有边界吸收特性; 当建立的是规则化的图模型时, 节点度 d_i 取相同的值 d_0 。对应的自吸收概率为

$$p_{ii} = h_i / (h_i + d_i) = \gamma_2 \times \text{rank}(\mathbf{G}_i) \times \sigma_i / (\gamma_2 \times \text{rank}(\mathbf{G}_i) \times \sigma_i + d_0 \times \bar{\sigma})$$

由于特征矩阵秩 $\text{rank}(\mathbf{G}_i)$ 的 4 种取值分别对应窗口内颜色的 4 种分布模型, 此时图像上各点的 p_{ii} 取值实质上是按照线-线、点-线、点-点及单点 4 种颜色模型先进行粗略地分段, 而后再用 $\sigma_i / \bar{\sigma}$ 进一步细化而得, 这样信息流将会根据图像的局部特性进行自适应地扩散。

σ_i 与 $\text{rank}(\mathbf{G}_i)$ 计算方法为: 首先对矩阵 \mathbf{G}_i 进行奇异值分解得 $\mathbf{G}_i = \mathbf{U} \Sigma \mathbf{V}^T$, 得到对角线上的奇异值为 $[\sigma_{i1} \quad \sigma_{i2} \quad \sigma_{i3} \quad \sigma_{i4}]$, 且 $\sigma_{i1} > \sigma_{i2} > \sigma_{i3} > \sigma_{i4}$, 则转换后的方差为 $\sigma_i = \sqrt{(\sigma_{i1}^2 + \sigma_{i2}^2 + \sigma_{i3}^2 + \sigma_{i4}^2) / 4}$, $\text{rank}(\mathbf{G}_i)$ 的

取值为

$$\text{rank}(\mathbf{G}_i) = \arg \max_{k \in \{1,2,3,4\}} \left[\frac{\sigma_{ik}}{\sigma_i} > t \right] \quad (12)$$

式中: t 为预先设定的阈值, 参考文献 [23] 取值为 0.002 5。式 (12) 表示转换空间中方差大于阈值 t 的维数。

4 单帧图片抠图

为了提高算法的快速性, 单帧抠图采用双层的形式。首先对图像进行 SLIC 超像素 [24] 分割, 构建基于超像素的图模型 \mathbf{G}_1 , 对初始的用户输入进行信息扩散得到高层抠图结果, 接着将高层结果作为低层抠图的输入标注信息, 在基于像素的图模型 \mathbf{G}_2 上进行扩散, 求得细化后的结果。本文在高层和低层都采用具有软约束的随机游走算法 SCRW, 二者的区别在于图模型的构造和控制矩阵 \mathbf{H} 的设计上。

在高层中, 图模型 $\mathbf{G}_1 = (\mathbf{V}_1, \mathbf{E}_1)$ 中的连接边有 3 种: 空间上相毗连的两个超像素点间; 共享一个边的超像素点间; 在颜色空间中用 FLANN 算法 [25] 寻找到的 n_k 个最相似的超像素点间。连接两节点的相似度计算公式为

$$w_{uv} = \exp\left(-\frac{\|c_u - c_v\|}{\delta^2}\right) \quad (13)$$

式中: c_u, c_v 为超像素节点 u, v 的平均 CIELAB 颜色值; δ 为控制相似度计算的常数, 文中 $\delta^2 = 0.1$ 。由于在全局图像的相似度计算中采用相同的 δ , 因此高层建立的图模型 $\mathbf{G}_1 = (\mathbf{V}_1, \mathbf{E}_1)$ 是未归一化的, 根据式 (11) 各节点的控制参数取相同的 γ_1 , 矩阵 $\mathbf{H} = \gamma_1 \mathbf{I}$ 。记某超像素 s 内的总像素数为 n_s , 用户标注的前景像素数为 n_s^f , 背景像素数为 n_s^b , 当 $n_s^f/n_s = 1, n_s^b/n_s = 1$ 时, 该超像素被定为前景、背景种子点; 当 $n_s^f/n_s < 1, n_s^b/n_s < 1$ 时, 该超像素是不完全标注, 取 $n_s^f/n_s, n_s^b/n_s$ 分别作为前景、背景的初始软约束条件, 运用软约束随机游走算法来计算超像素属于前景和背景的概率, 从而得到高层抠图结果。

在低层中, 图模型按照传统的 CF 算法 [10] 建立, 相似度函数为

$$w_{i,j} = \sum_{q|(i,j) \in \Omega_q} \frac{1}{|\Omega_q|} \left(1 + (\mathbf{I}_i - \mu_q)^T \left(\sum_q + \frac{\varepsilon}{|\Omega_q|} \Lambda_3 \right)^{-1} (\mathbf{I}_j - \mu_q) \right)$$

式中: Ω_q 为围绕像素点 q 的 3×3 窗口; μ_q, Σ_q 为窗口内颜色的均值和方差; $|\Omega_q|$ 为窗口像素数; ε 为规则化因子, 取 10^{-5} 。由以上分析可知, 此时建立的是归一化图模型。根据式 (12) 计算各节点窗口内的颜色特征向量的秩与方差, 再代入式 (11)

得到各节点对应的输入控制参数。低层的种子点包含原始用户输入的标注与高层计算中新增加的种子点, 新增前背景种子点的阈值取为

$$\begin{cases} t_f = \frac{\text{mean}(\alpha_s) + 1.2 \times \max(\alpha_s)}{2} \\ t_b = 0.5 \times \text{mean}(\alpha_s) \end{cases}$$

式中 α_s 为高层基于超像素计算出来的 α 值。当 $\alpha_i > t_f$ 时, 则点 i 取为新的前景种子点; 当 $\alpha_i < t_b$ 时, 则点 i 取为新的背景种子点; 其他点则将高层传递下来的 α 值作为软约束条件, 引导随机游走的相似扩散过程, 从而得到低层的抠图结果。通过新增加前景和背景种子点, 在低层的 SCRW 抠图中, 未知点的数目减少, 因而算法的运行时间降低。

5 视频抠图

由于视频抠图处理数据量大, 一般无法对每帧图像进行标注, 但图像序列间具有连续性与相似性, 充分利用图像的帧间信息可以获得单张图像不具备的特征。本文在单帧双层 SCRW 算法的基础上, 采用软硬两种约束相结合的方式视频抠图。图 1 为视频抠图的示意图: 左侧为输入的第 $i-1$ 帧、第 i 帧图像及相应的三分图区域 (背景 B, 前景 F, 未知区域 U)。连续两帧图像间的信息传导有两种: 光流映射与流形最近邻映射。

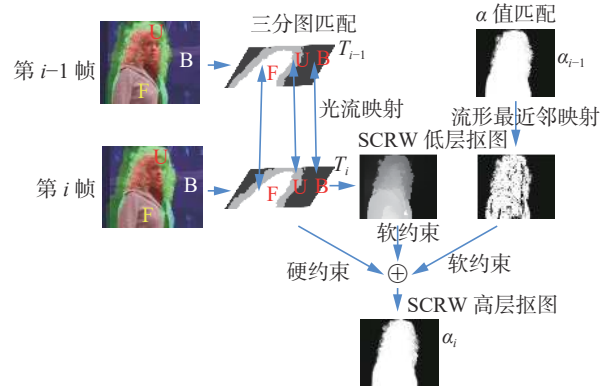


图 1 SCRW 视频抠图示意

Fig. 1 Schematic of SCRW video matting process

首先计算图像的前向与后向运动向量 [26-27] $[u_b \ v_b \ u_f \ v_f]$, 将前一帧的三分图 T_{i-1} 按照前向光流映射到当前帧, 并对其进行形态学操作去除部分杂点, 确保新产生的三分图 T_i 的准确性。而后对图像进行超像素划分进行高层 SCRW 运算, 将已知区域向未知区域扩散得到初步抠图结果 α_b 。与式 (13) 不同, 此时具有边连接的两超像素相似度的计算中包含了前向、后向光流场向量:

$$w_{kl} = \exp\left(-\frac{\|\mathbf{F}_k - \mathbf{F}_l\|}{\delta^2}\right)$$

式中: \mathbf{F}_k 为颜色特征; \mathbf{F}_l 为运动向量特征, 取值为

$$\mathbf{F} = [l \ h \ v \ \lambda_f u_b \ \lambda_f v_b \ \lambda_f u_f \ \lambda_f v_f] \quad (14)$$

式中 λ_f 控制颜色特征与运动特征的比重,这意味着当图像中的点拥有相近的颜色及运动时,其对应的 α 值也相近。在三分图标注信息下取得的高层SCRW结果 α_h 较为粗糙,需要低层的进一步精细化处理,除了将高层结果作为软约束外,本文还用FLANN算法^[25]为当前帧中的点在前一帧寻找其最优近邻,最近邻的搜索特征为

$$\mathbf{X} = [r \ g \ b \ \lambda_{s,x} \ \lambda_{s,y}] \quad (15)$$

上述FLANN搜索特征中未加入光流向量是因为不同帧间的光流特征不具有可比性。与传统视频抠图要求三分图密实围绕前景物体、宽度较窄不同,本文只需要勾画前景与背景的大致区域,即只需要前背景的稀疏输入,因而未知区域的范围较宽。然而未知区域通常包含复杂的图像细节,如毛发、孔洞等,且由于图像运动造成的前景物体遮挡、新增前景等原因,当前帧未知区域中的点与前一帧的流形匹配点有可能差别较大,置信度低。设 $\mathbf{M} = [m_i]_{n \times 1}$ 为当前帧中的点与其前帧最优匹配点间的欧式距离,对向量 \mathbf{M} 取阈值 T_d ,丢弃距离大于 T_d 的点则得到流形约束图 α_m 。此时除了式(11)的两种输入控制参数外,流形约束 α_m 对应的输入控制矩阵设计为

$$h_i = \gamma_3 \times \left(1 - \frac{m_i}{\max(m_i)}\right) \quad (16)$$

式中 γ_3 为可控常量。本文除了通过光流匹配的三分图 T 传递,还有通过流形匹配的 α 值信息传递,且高层的超像素相似传播包含了流形特征。而大多数传统算法只是在三分图的产生过程中用到了光流信息,因此本文算法能够取得更为良好的时空一致性的效果。本文视频抠图的基本步骤为:

- 1) 用光流法计算视频序列中每一帧的前向、后向光流运动向量 (u_b, v_b, u_f, v_f) ;
- 2) 用前向光流 (u_f, v_f) 将前帧三分图 T_{i-1} 匹配到当前帧,得到三分图 T_i ;
- 3) 以 T_i 为硬约束条件,对当前帧进行高层SCRW相似扩散得到 α_h ;
- 4) 通过流形最近邻映射将前帧 α 值匹配到当前帧,并去除置信度低的点,得到前帧 α 值约束 α_m ;
- 5) 综合三分图硬约束 T_i 与软约束 α_h 、 α_m ,及式(11)、式(16)的输入控制矩阵,引导底层SCRW相似扩散,得到最后的结果 α ;
- 6) 将 $\alpha_i > 0.8$ 、 $\alpha_i < 0.2$ 作为确定的前景与背景,并对其未知区域向外膨胀 n_c 个像素,重新产生向后传递的三分图 T'_i ;

7) 判断视频是否处理完毕或到预定处理帧数,否则转到2)。

6 实验结果

实验的运行环境为: Intel Core i3 双核 3.3 GHz CPU, 编程环境为 Matlab 2016。对比算法有CF^[10]、LB^[11]、KNN^[12]抠图算法。由于CF算法在层数大于5时,程序无法正常运行,因此取层数level=1及level=5两种水平下的结果与本文算法进行对比,文中超像素数取1000。实验中定量分析图片源于抠图标准网站 <http://www.alphamattin.com>^[28]。

6.1 单帧抠图实验结果与分析

1) 参数 γ_1 、 γ_2

为了验证参数 γ_1 、 γ_2 对信息传播的影响,此处忽略各节点的差异性,将上下两层的 \mathbf{H} 矩阵取为单位矩阵。如图2所示,左上图像为带笔画标注的图像,白色画线为前景标注,黑色画线为背景标注。右上为该图在不同 γ_1 、 γ_2 参数下的MSE误差,下方为运行结果。从图2中可以看出,随着参数 γ_1 、 γ_2 的降低,各节点的自吸收概率降低,已标注区域的信息有更大的概率达到未知区域,信息流有更远的传播距离,因而距离已知区域较远的点也能得到引导信息,从而使误差MSE值降低。经过多次实验,选择参数 γ_1 、 γ_2 的值为 10^{-2} 、 10^{-3} ,并将其应用到后续的实验。

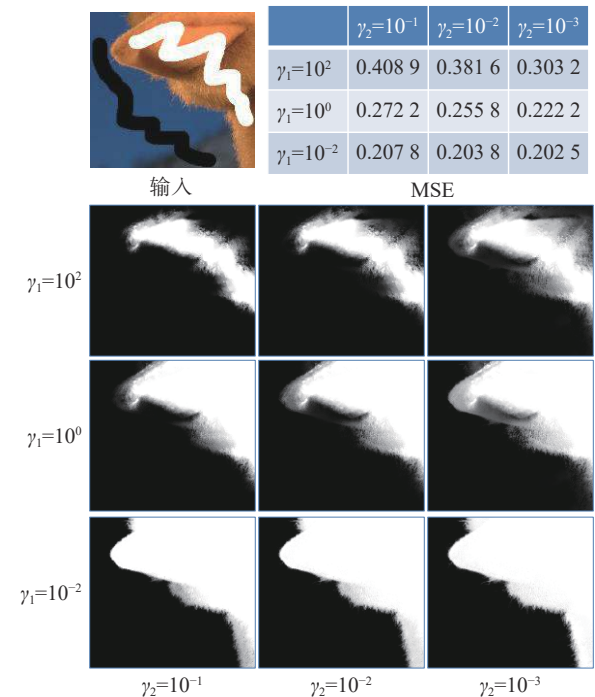


图2 全局输入控制参数 γ_1 、 γ_2 的影响

Fig. 2 Influence of global input control parameters γ_1 , γ_2

2) 容错性

图3是不完全正确标注情况下,传统CF算法

和本文所提 SCRW 算法的容错性比较。在图 3(b) 中,白色画线为前景标注,黑色画线为背景标注,中间圆圈区域内,用户错误地将中间背景布标注为前景。图 3(c) 的传统 CF 抠图算法,因为采用 λ 为大值的硬性约束,输出严格遵从输入,最终结果仍然将中间的背景布抠选了出来;而在图 3(d) 中,因为 SCRW 采用 λ 为小值的软性约束,各节点局部自适应的 h_i 设计使得输入信息能够根据图像特征进行自适应扩散,可对一些非正确的输入进行校正,使其符合图像内容,因此本文算法的容错性更强。



图3 容错性比较

Fig. 3 Comparison of fault tolerances

3) 定量比较

在图 4(a) 中,三分图 1 是原始三分图,三分图 2~三分图 5 是原三分图的未知区域向外扩展 10、15、20、25 个像素点而得,笔画式标注则是手动标注(白色前景,黑色背景),图 4(a) 的用户标注信息从左至右依次减少。图 4(b)、(c) 为各算法在不同的用户标注下的 MSE 误差及运算时间。由图 4(b) 知,在三分图比较紧凑的情况下,未知区域中的点距离已知区域近,不需远距离的信息传播就能获得引导信息,CF、LB 算法的 MSE 误差反而比较小;但随着用户标注区域的减少,CF、LB 算法由于缺乏远距传播,未知区域中的部分点得不到引导信息,产生了较大的误差,虽然 CF 算法随着降采样水平的增高,也具有一定的远距传播能力,但其过强的输入控制使高层中不够准确的结果在低层中产生误扩散,故其 MSE 误差在高 level 的情况下反而更大;KNN 算法虽采用非局部近邻,但其流形近邻的搜索特征包含空间信息,传播距离也有限;而本文算法的高层远距离传播及低层细节恢复能力使得算法的整体误差最小,尤其是在用户输入信息不够充分的情况下能够取得更好的结果。在图 4(c) 中,CF 算法在 level=5 时的运算时间最短,但其相应的误差最大,是以牺牲准确度来降低运算时间;除此之外,随着未知区域的增加,各算法的运算时间出现不同程度的增加,但本文算法的增量最小,运算时间基本在 10 s 左右。

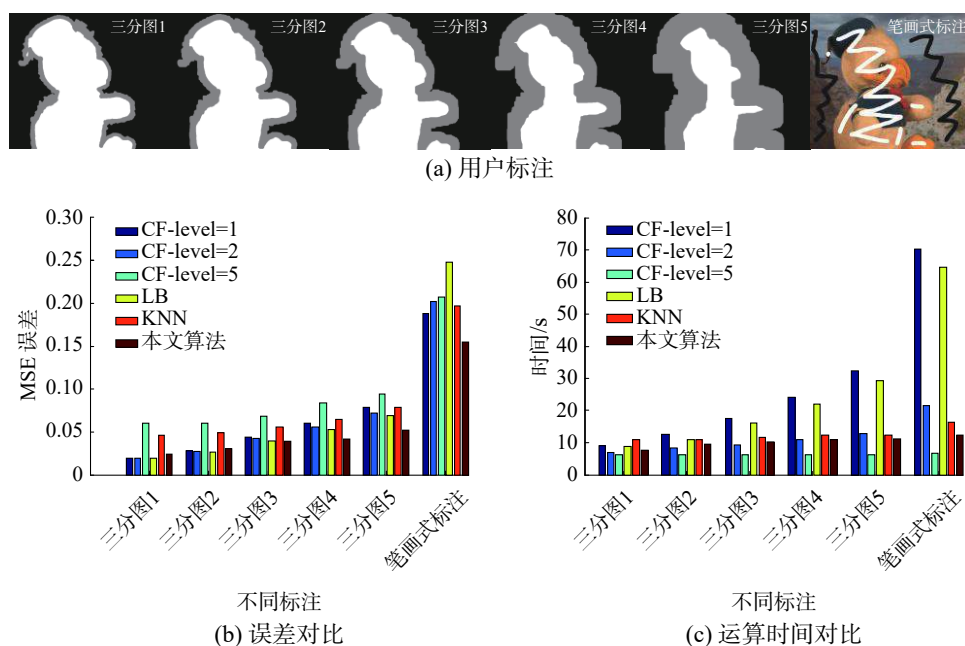


图4 不同用户输入下各算法准确率及运算时间对比

Fig. 4 Comparison of algorithm accuracy and operation time with different user inputs

6.2 视频抠图实验结果与分析

实验中特征权重参数 $\lambda_f = 0.001$ 、 $\lambda_s = 1$, 膨胀像素数 $n_e = 50$, 阈值 $T_d = 0.002$, 输入控制参数 $\gamma_3 = 0.001$ 。式 (14)、式 (15) 中的颜色特征需经过归一化处理, 视频数据来源于文献 [29]。

如图 5 所示, 是 Amira 视频第 79 帧到第 82 帧的各算法运行结果。图 5(b) 的三分图为前帧通过光流匹配得到的, 图 5(a) 前背景边界线与图 5(b) 中的三分图对应。由图 5(a) 知, 随着前景图像的运动, 匹配的三分图虽然大致划分了前景背景区域, 但并没有很好地贴近前景边界, 虽然图 5(d)

中的高层软约束 SCRW 能够增加输入信息的远距传播能力, 但因前景物体与背景的颜色存在一定的相似度, 因此 Amira 图像的灰色外套及眼部没有完全地被抠选出来 (图中箭头所示)。在图 5(e) 中, 由于 CF 算法的局部窗口作用会使得算法在确定的前景和背景间平滑过渡, 未知区域中的前景物体出现了半透明的抠图结果。本文所提出的带流形匹配的 SCRW 算法由于存在三分图和前帧 α 值传递的软约束, 能在较大未知区域中提供引导信息, 因此能够取得时空更加一致的抠图结果。

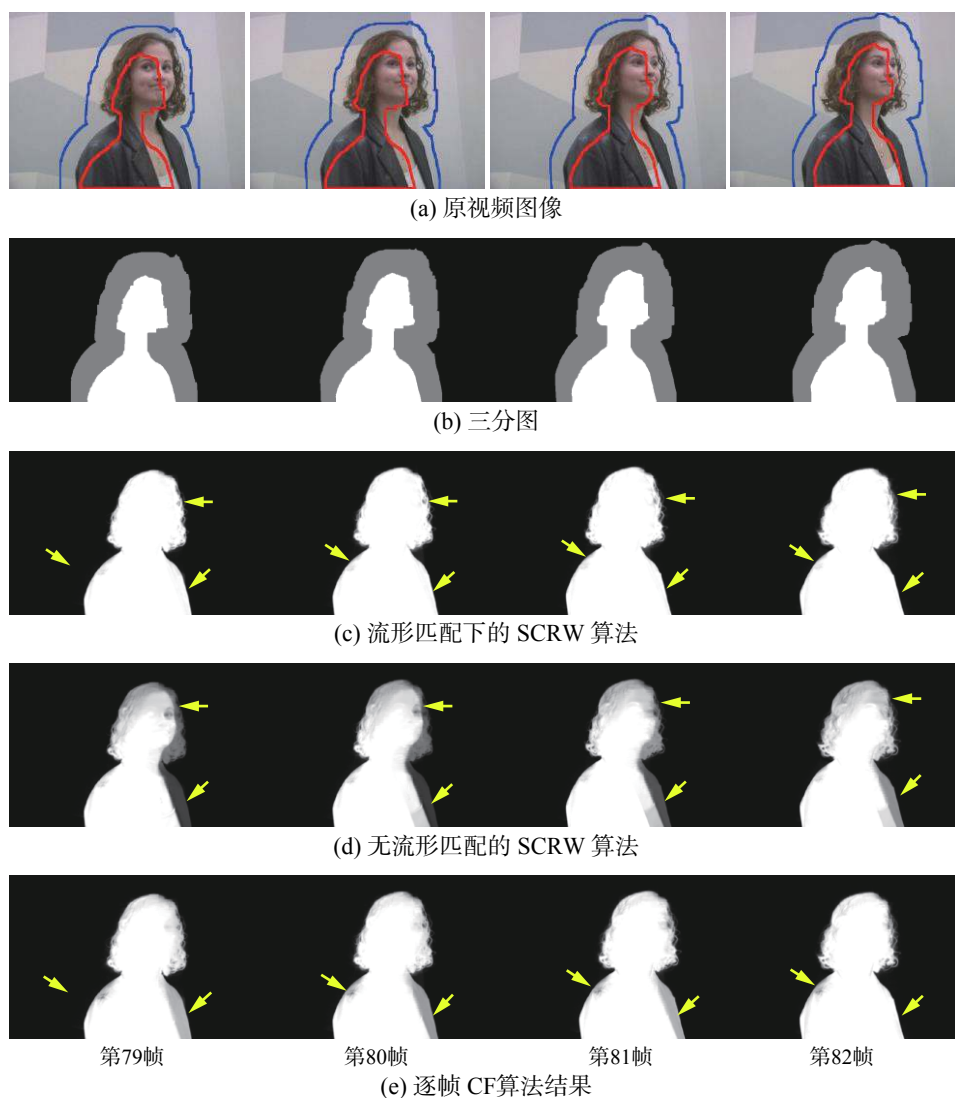


图 5 Amira 视频序列抠图结果

Fig. 5 Matting results of Amira video sequence

7 结束语

本文针对抠图算法标注准确性问题, 根据带约束随机游走算法信息流的传播特性, 提出了一种带软性约束的随机游走算法, 并将其应用到单帧图像抠图和视频抠图中。实验结果表明, 输入

控制参数对标注信息的扩散距离具有直接的影响, 软性约束下随机游走具有更加优良的容错特性, 所提算法避免了视频抠图中获取三分图标注的繁杂计算, 为提高抠图算法的精度提供了新的思考方向。但本文算法在大尺寸图像和视频上运

行缓慢,对边界模糊图像的处理效果不够理想,如何进一步提高算法的快速性和复杂图像的处理能力仍将是未来努力的方向。

参考文献:

- [1] YAO Guilin, ZHAO Zhijie, LIU Shaohui. A comprehensive survey on sampling-based image matting[J]. *Computer graphics forum*, 2017, 36(8): 613–628.
- [2] ZHU Qingsong, SHAO Ling, LI Xuelong, et al. Targeting accurate object extraction from an image: a comprehensive study of natural image matting[J]. *IEEE transactions on neural networks and learning systems*, 2015, 26(2): 185–207.
- [3] KARACAN L, ERDEM A, ERDEM E. Alpha matting with KL-divergence-based sparse sampling[J]. *IEEE transactions on image processing*, 2017, 26(9): 4523–4536.
- [4] XU Ning, PRICE B, COHEN S, et al. Deep image matting[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA, 2017: 311–320.
- [5] SHI Yongfang, AU O C, PANG Jiahao, et al. Color clustering matting[C]//Proceedings of 2013 IEEE International Conference on Multimedia and Expo. San Jose, USA, 2013: 1–6.
- [6] JIN Meiguang, KIM B K, SONG W J. Adaptive propagation-based color-sampling for alpha matting[J]. *IEEE transactions on circuits and systems for video technology*, 2014, 24(7): 1101–1110.
- [7] HE Bei, WANG Guijin, ZHANG Cha. Iterative transductive learning for automatic image segmentation and matting with RGB-D data[J]. *Journal of visual communication and image representation*, 2014, 25(5): 1031–1043.
- [8] JOHNSON J, VARNOUSFADERANI E S. Sparse coding for alpha matting[J]. *IEEE transactions on image processing*, 2016, 25(7): 3032–3043.
- [9] LI Xuelong, LIU Kang, DONG Yongsheng, et al. Patch alignment manifold matting[J]. *IEEE transactions on neural networks and learning systems*, 2018, 29(7): 3214–3226.
- [10] LEVIN A, LISCHINSKI D, WEISS Y. A closed-form solution to natural image matting[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2008, 30(2): 228–242.
- [11] ZHENG Yuanjie, KAMBHAMETTU C. Learning based digital matting[C]//Proceedings of the 12th International Conference on Computer Vision (ICCV). Kyoto, Japan, 2009: 889–896.
- [12] CHEN Qifeng, LI Dingzeyu, TANG C K. KNN matting[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(9): 2175–2188.
- [13] TSENG C Y, WANG S J. Learning-based hierarchical graph for unsupervised matting and foreground estimation[J]. *IEEE transactions on image processing*, 2014, 23(12): 4941–4953.
- [14] GONG Minglun, QIAN Yiming, CHENG Li. Integrated foreground segmentation and boundary matting for live videos[J]. *IEEE transactions on image processing*, 2015, 24(4): 1356–1370.
- [15] LEE S Y, YOON J C, LEE I K. Temporally coherent video matting[J]. *Graphical models*, 2010, 72(3): 25–33.
- [16] SHAHRIAN E, PRICE B, COHEN S, et al. Temporally coherent and spatially accurate video matting[J]. *Computer graphics forum*, 2014, 33(2): 381–390.
- [17] LI Dingzeyu, CHEN Qifeng, TANG C K. Motion-aware KNN Laplacian for video matting[C]//Proceedings of 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia, 2013: 3599–3606.
- [18] SINDEEV M, KONUSHIN A, ROTHER C. Alpha-flow for video matting[C]//Proceedings of the 11th Asian Conference on Computer Vision. Daejeon, Korea, 2012: 438–452.
- [19] CHO D, KIM S, TAI Y W, et al. Automatic trimap generation and consistent matting for light-field images[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(8): 1504–1517.
- [20] CHO H W, CHO Y R, SONG W J, et al. Image matting for automatic target recognition[J]. *IEEE transactions on aerospace and electronic systems*, 2017, 53(5): 2233–2250.
- [21] GRADY L. Random walks for image segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(10): 1768–1783.
- [22] WU Xiaoming, LI Zhenguo, SO A M C, et al. Learning with partially absorbing random walks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS). Lake Tahoe, USA, 2012: 3077–3085.
- [23] SINGARAJU D, ROTHER C, RHEMANN C. New appearance models for natural image matting[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami Beach, USA, 2009: 659–666.
- [24] ACHANTA R, SHAJI A, SMITH K, et al. SLIC super-

- pixels compared to state-of-the-art superpixel methods[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(11): 2274–2282.
- [25] MUJA M, LOWE D G. Fast approximate nearest neighbors with automatic algorithm configuration[C]//Proceedings of the International Conference on Computer Vision Theory and Applications. Lisbon, Portugal, 2009: 331–340.
- [26] BAKER S, SCHARSTEIN D, LEWIS J P, et al. A database and evaluation methodology for optical flow[J]. *International journal of computer vision*, 2011, 92(1): 1–31.
- [27] BROX T, BRUHN A, PAPENBERG N, et al. High accuracy optical flow estimation based on a theory for warping[C]//Proceedings of the 8th European Conference on Computer Vision (ECCV). Prague, Czech Republic, 2004: 25–36.
- [28] RHEMANN C, ROTHER C, WANG Jue, et al. A perceptually motivated online benchmark for image matting[EB/OL].(2009-06)[2018-09-11] <http://www.alphamatting.com/>.
- [29] CHUANG Y Y, AGARWALA A, CURLESS B, et al. Video matting of complex scenes[EB/OL]. (2002-07)[2018-09-11] <http://grail.cs.washington.edu/projects/digital-matting/video-matting/>.

作者简介:



陈秋凤, 女, 1983 年生, 讲师, 博士, 主要研究方向为图像处理、智能控制。发表学术论文 12 篇。



申群太, 男, 1944 年生, 教授, 博士生导师, 主要研究方向为人工智能、工业控制。曾获湖南省科技进步三等奖及国家科技进步二等奖。发表学术论文 200 余篇。