

一种基于局部学习的自然图像景物提取方法^{*}

彭宏京¹⁺, 陈松灿², 张道强²

¹(南京工业大学 计算机科学系,江苏 南京 210009)

²(南京航空航天大学 计算机科学与工程系,江苏 南京 210016)

A Local Learning Approach for Natural Image Matting

PENG Hong-Jing¹⁺, CHEN Song-Can², ZHANG Dao-Qiang²

¹(Department of Computer Science, Nanjing University of Technology, Nanjing 210009, China)

²(Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

+ Corresponding author: E-mail: hjpeng@njut.edu.cn

Peng HJ, Chen SC, Zhang DQ. A local learning approach for natural image matting. *Journal of Software*, 2009,20(4):834-844. <http://www.jos.org.cn/1000-9825/3439.htm>

Abstract: A scheme of categorizing Laplacians is introduced in this paper based on the computation times of similarity weights for each pair of adjacent data points. It is also theoretically proven that the Laplacian construction with multiple computations of similarity weights for each pair of adjacent points can better capture the local intrinsic structure of data than those methods with only one or two such computations. A novel Laplacian construction method is then proposed, which is more suitable for natural image matting task. In this method, all the different similarity weights for any pair of adjacent pixels are reconstructed by using a local linear model in the neighborhoods they fall into. By combining the user-provided constraints which specify some pixels as foreground or background, a quadratic objective function for matting based on semi-supervised learning is formed. When estimating the colors of unknown pixels by sampling foreground and background colors, this optimization problem is reformulated and solved in an iterative manner. What's more, this iterative scheme can also be successfully generalized and applied into other previously constructed Laplacians for image matting tasks with only sparse label scribbles. Both the theoretical analysis and experimental results validate that the proposed Laplacian construction approach can better capture the intrinsic structure between image pixels, and can propagate the finer ingredients of an image foreground and background rather than just their labels, and thus the mattes of higher quality are obtained.

Key words: matting; semi-supervised learning; local learning; Laplacian regularization; quadratic optimization

摘要: 引入一种按邻点对的相似性权值计算次数来归类 Laplacian 的思想,并从理论上证明了包含多次相似性权值计算的 Laplacian 构造比只计算一次或两次相似性权值的 Laplacian 构造更能精细地刻画数据局部几何结构.据此提出了一种新的更能胜任自然图像景物提取任务的 Laplacian 构造方法.该方法通过任意一对相邻像素在不同局部邻域内建立一个线性学习模型来重构不同的相似性权值.结合用户提供的部分前、背景标记约束,导出求解景物提取的半监督二次优化目标函数.当考虑通过对前、背景抽样来估计未知像素的颜色值时,优化目标可以迭代求

* Supported by the National Natural Science Foundations of China under Grant Nos.60505004, 60773061 (国家自然科学基金)

Received 2008-01-18; Accepted 2008-08-13

解.更有意义的是,该迭代方法可以成功地将原来构造的其他 Laplacian 推广应用于只提供稀疏指示条带的景物提取问题中.理论分析与实验结果均证实,所构造的 Laplacian 能够更充分地表达图像像素间的内在结构,能以更精细的方式约束传播前、背景的成分比例而不仅仅是标号,从而获得更优的景物提取效果.

关键词: 景物提取;半监督学习;局部学习;拉普拉斯正则化;二次优化

中图法分类号: TP181 **文献标识码:** A

景物提取(matting)是一种把图像的前景部分从背景中分离出来的技术,广泛应用于媒体特效制作、娱乐和虚拟现实等各种领域,是一种不可或缺的图像编辑工具.一旦景物被分离出来,就容易与其他背景图像合成而获得新的图像.景物提取问题将一幅输入图像 C 建模为前景图像 F 和背景图像 B 的凸组合,即满足如下图像合成方程:

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

这里, $\alpha(p)$ 是指任意像素 p 的混合参数(也称为透明度 α).因而从背景中分离景物的问题就变成参数估计问题.对于自然图像来说,其挑战在于对背景不作任何约束,导致了一个严格的欠适定问题.

现有的自然图像景物提取技术,依据它们如何利用图像统计和近邻像素的相关性,可粗略地分为基于抽样和基于传播的两类方法^[1].基于抽样的方法假设任何未知的待估像素的前、背景颜色值均可直接由用户提供的近邻像素来估计.贝叶斯方法(Bayesian matting)^[2]是该类方法的典型代表,它利用了图像的颜色统计,沿着已估前景和背景区域的边界逐个像素地进行分布估计,通过最大化似然获得问题的解.贝叶斯方法及一些较早期的景物提取方法^[3]需要事先在提取目标的边界给出已知前、背景和未知待估区域的精细划分(称为 trimap),在提取复杂景物时,这种要求无疑对用户是一种负担,有时甚至无法进行.信任传播方法(belief propagation,简称 BP)^[4]构造了一个迫使前、背景符合局部统计约束的能量最小化问题,用户只需提供稀疏的前、背景指示条带,利用高代价的信任传播非线性迭代优化过程来估计每个像素的 α 值.Easy Matting^[5]基于待估区域的马尔可夫随机场建模,通过由粗到精选代传播求解能量方程,导出对连续 α 值的估计,但由于其构造的光滑约束项缺乏保持边缘细节的能力,因此在稀疏条带不足以充分代表前、背景颜色分布的情况下,必须辅助局部抽样过程来获得足够的样本再次对结果优化.与之相对照,基于传播的方法并不显式地估计前、背景的颜色值,而是基于前、背景满足局部光滑性约束和未知像素邻域图像假设为恒定或线性变化的前提下,使得混合参数求解过程中能够容易地消除前、背景成分,从而可以获得 α 的闭解.如何让信息在像素之间可靠传播,是该类方法能否成功提取景物目标的关键,而信息传播的可靠性程度又取决于获取近邻像素相关性的准确度.这种局部结构的刻画在图像处理 and 模式识别等领域具有普遍意义.Sun 等人^[6]在给定的 trimap 划分下假设前、背景颜色分量光滑,由 α 的梯度来反映图像的局部变化,从而导出求解 α 的泊松(Poisson)方程;随机步行(random walk)方法^[7]和闭解(closed-form)方法^[8]都是通过最小化二次代价函数来传播用户提供的前、背景约束以完成整幅图像的前景目标和背景图像的分离.Robust Matting^[1]试图结合上述两类技术,通过在抽样阶段选择具有高信任度的样本估计出一个初始的 α 值,然后由随机步行最优化一个图标记问题来进一步改进结果.因此,该方法对初始提供的标记样本有较强的依赖性.如果手工提供的初始前、背景指示条带分布稀疏,则难以保证抽样估计的初始 α 值是合理的.Levin 等人^[9]对文献[8]中构造的 Laplacian 矩阵的谱进行了进一步分析,提出了谱景物提取方法.文献[10]采用像素属于前、背景的似然梯度为权重计算每个像素到用户提供的指示条带的加权距离,再由这些距离判断每个像素隶属于哪个条带的方法,能够快速提取前景目标.

上述多数提取自然图像景物的方法都直接或间接地与 Laplacian 的构造有关,或最终归结为一个能量最小化问题.它们的差异主要表现在采用了不同的 Laplacian 光滑约束项或者结合了不同的优化求解技术.然而,对于采用不同的 Laplacian 导致不同的求解结果,以及应该构造怎样的 Laplacian 才能获得理想的结果等问题,至今仍然缺乏理论分析.

针对上述问题,本文将用户提供部分前、背景指示约束下提取图像目标景物的问题纳入到一类半监督直推式学习框架之中.首次提出按相似性权值计算次数对 Laplacian 构造方式进行分类的思想,将目前的局部学习技

术^[11-16]所采用的几种邻域相关性描述方式进行系统化;另外,基于任何一对相邻像素在不同邻域所贡献的相似性程度也不同的直观认识,提出了一种新的、包含多次相似性权值计算的 Laplacian 构造方法;进一步将上述所构造的 Laplacian 与抽样技术相结合,提出一种迭代求精的景物提取求解方法.

1 二次目标函数及其 Laplacian 正则化矩阵的构造和分类

给定一个点集 $X=\{x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_n\} \subset R^d$ 和一个标号集 $Y=\{1, 2, \dots, c\}$, 前 l 个点 $x_i (i \leq l)$ 标记为 $y_i \in Y$, 而剩下的点 $x_u (l+1 \leq u \leq n)$ 是未标记的点, 考虑一类半监督分类方法, 它们可以表示成如下的二次优化问题^[16]:

$$J = \frac{1}{2} f^T L f + (f - y)^T E (f - y) \quad (2)$$

这里 $y=[y_1, \dots, y_l, 0, \dots, 0]^T \in R^n$, $L \in R^{n \times n}$ 是由点集构造的正则化矩阵, 求解目标 $f=[f_1, f_2, \dots, f_n]^T \in R^n$ 的意义依问题而定: 对于分类和图像分割 f 是指类别标号; 而对于景物提取问题 f 是指 α 值, 等等. $E \in R^{n \times n}$ 是对角矩阵, 其对角元素的选取, 依据希望对求解结果实行怎样的约束而定.

半监督学习问题的关键是基于先验一致性假设: 即相邻点最可能有相同的标号; 在同样结构上(流形或聚类)的两个数据点更可能有相同的标号. 各种半监督学习算法的主要差异主要体现在实现先验一致性假设的方式有所不同. 针对半监督二次优化问题(2), 如何构造正则化矩阵 L 是关键. 下面我们按照在构造 L 时任一邻点对的相似性权值计算次数, 将 Laplacian 矩阵划分为 3 种类型.

定义 1. 任一邻点对相似性权值计算一次、两次和多次所构造的 Laplacian 分别称为单相关性 Laplacian、重相关性 Laplacian 和多相关性 Laplacian.

图 1 示意性地比较了上述 3 种构造的 Laplacian 在描述两个相邻点之间相似性时的差异. 网格上的点表示图的顶点, i 和 j 表示一对相邻点, 当邻点对 (i, j) 出现在以某个点为中心的邻域而需要重新计算它们的权值时, 以粗线框表示该邻域窗口(窗口大小为 3×3). 这里用 $w_{ij}(k)$ 表示以 k 为中心的邻域窗口中邻点对 (i, j) 的相似性权值.

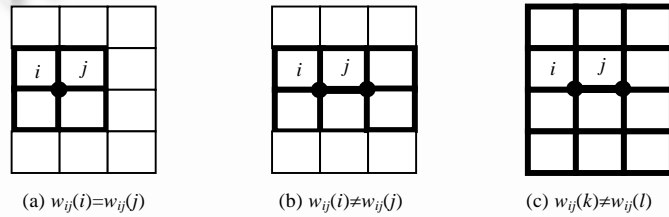


Fig.1 Geometry computations of local neighborhood vary with Laplacian construction methods

图 1 不同的 Laplacian 构造方法对应不同的局部邻域几何计算

单相关性 Laplacian 对应图 1(a) 的表示, 无论以 i 还是以 j 为中心的邻域, w_{ij} 只需计算一次, 即 $w_{ij}(i) = w_{ij}(j)$. 最典型的单相关性 Laplacian 构造方式是图 Laplacian^[12,13], 定义为 $L = D - W$. 其邻点对的相似性权重是由具有对称性的高斯核函数来计算的.

重相关性 Laplacian 对应图 1(b) 的表示, 当以 i 或 j 为中心的邻域重构权重时, w_{ij} 计算两次, 并且通常 $w_{ij}(i) \neq w_{ij}(j)$. 文献[11,14,15]构造了这样的重相关性 Laplacian: $L = (I - W)^T (I - W)$, 其权重矩阵是通过以每个数据点为中心的局部邻域线性重构得到的.

多相关性 Laplacian 对应图 1(c) 的表示, 只要相邻顶点对 (i, j) 出现在任何以 k 为中心的邻域中就包含了一次 w_{ij} 的计算, 并且对于 $k \neq l$, 一般有 $w_{ij}(k) \neq w_{ij}(l)$. 文献[8]构造的 Matting Laplacian 是多相关性 Laplacian 的一个例子, 具体构造方法参见文献[8].

命题 1. 多相关性 Laplacian 比单相关性 Laplacian 和重相关性 Laplacian 能够更有效地约束类别成分的传播.

证明: 根据先验一致性假设, 数据点 j 向邻点 i 传播类别信息的数量取决于它们之间的相似性程度. 假设任一数据点 j 的类别信息 h_j 进一步细分为若干成分, 并由这些成分加权求和来表示 $h_j = t_{j1}h^1 + t_{j2}h^2 + \dots + t_{jK}h^K$, 系数 t_{jK} 是

指类别信息成分 h^k 在 h_j 中所占的比例. 逐个成分地按照相似性关系考虑数据点 j 向数据点 i 传播类别信息, 那么数据点 i 从数据点 j 获得的类别信息可以表示为按成分相似性程度的加权组合形式, 即

$$h_i^* = w_{ij}^1 t_{j1} h^1 + w_{ij}^2 t_{j2} h^2 + \dots + w_{ij}^K t_{jK} h^K$$

w_{ij}^k 表示 (i, j) 数据点对在成分分量 h^k 上的相似性权. 不失一般性, 进一步假设在以 $k (k \neq j)$ 为中心的邻域中, 认为某个标号成分 h^k 占主导, $h^k \gg h^l (l=1, \dots, K, l \neq k)$, 则数据点对 (i, j) 在分量 h^k 上的相似性权重可以近似地用点对 (i, j) 落在以 k 为中心的邻域中的相似性权重 $w_{ij}(k)$ 来代替, 即 $w_{ij}^k \approx w_{ij}(k)$, 因此,

$$h_i^* \approx w_{ij}(1) t_{j1} h^1 + w_{ij}(2) t_{j2} h^2 + \dots + w_{ij}(K) t_{jK} h^K.$$

这就意味着类别成分依据相邻点对在不同邻域相似性权值的比例来传播. 那么, 3 种 Laplacian 构造方法约束类别信息的传播分别具有如下特点:

对于单相关性 Laplacian, 因为点对 (i, j) 之间的相似性权值只有一个, 它代表了所有成分的相似性关系, 因此, 向点 i 传播的类别信息为 $h_i^* = w_{ij}(t_{j1} h^1 + t_{j2} h^2 + \dots + t_{jK} h^K)$, 对各类别成分的贡献未作任何区分. 约束信息传播的方式相当于各向同性均匀扩散, 不具有局部保持性;

对于重相关性 Laplacian, 因为点对 (i, j) 之间的相似性权值有两个, 并且通常 $w_{ij}(i) \neq w_{ij}(j)$, 因此, 向点 i 传播的类别信息为 $h_i^* = w_{ij}(1) t_{j1} h^1 + w_{ij}(2) t_{j2} h^2 + \dots + w_{ij}(K) t_{jK} h^K$, 但 $w_{ij}(i) = w_{ij}(i)$ 或 $w_{ij}(j) = w_{ij}(j)$. 即只粗略地区分各类别成分的贡献;

对于多相关性 Laplacian, 因为点对 (i, j) 之间的相似性权值有多个, 并且通常它们互不相同, 因此, 向点 i 传播的类别信息为 $h_i^* = w_{ij}(1) t_{j1} h^1 + w_{ij}(2) t_{j2} h^2 + \dots + w_{ij}(K) t_{jK} h^K$, 对任意 $l \neq k, w_{ij}(l) \neq w_{ij}(k)$. 这就意味着该构造方法能够更精细地区分所有标号成分的贡献, 并能更好地在相邻点间依据它们的不同局部相关性来传播类别成分. \square

显然, 只有当在以 j 为中心的邻域中某个成分占绝对优势时, 上述 3 种情形才基本相当.

2 基于局部学习的景物提取方法

在用户提供部分前、背景标记指示下进行自然图像景物提取的过程, 实质上属于半监督学习问题. 在下面的讨论中, 上节中的数据点集即指图像像素. 根据第 1 节的分析, 需要构造符合图 1(c) 特点的多相关性 Laplacian. 为此, 我们建立局部线性学习模型来直接重构图 1(c) 中的那些权重. 一旦 Laplacian 构造完成, 结合用户提供的部分前、背景指示约束, 景物提取问题最终变成如公式(2)所示的二次优化求解问题. 对照公式(2), 本文从景物提取的局部学习一般模型出发, 分别介绍多相关性 Laplacian 正则化项构造方法和数据拟合约束项的构造方法.

2.1 景物提取的局部学习统一模型

为了求解一幅图像 C 的前景 F 和背景 B 的混合参数 α , 假设可以建立一个局部学习模型来预测. 定义如下二次目标函数:

$$J_s = \sum_{j=1}^n \sum_{i \in N_j} (\alpha_i - o_{ij})^2 \quad (3)$$

其中, o_{ij} 表示第 i 个像素在以 j 为中心的邻域 N_j 中通过某个局部学习模型预测的 α 输出值, 本文取 3×3 的邻域窗口, n 指图像的像素数. 改写公式(3)为如下矩阵形式:

$$J_s = \sum_m \|\alpha_m - o_m\|^2 \quad (4)$$

这里, $o_m = [o_{1m} \ o_{2m} \ \dots \ o_{|N_m| m}]^T$, $\alpha_m = [\alpha_1^{(m)} \ \alpha_2^{(m)} \ \dots \ \alpha_{|N_m|}^{(m)}]^T$ 是第 m 个窗口中像素的 α 构成的向量, $|N_m|$ 是指第 m 个小窗口的像素数, m 是图像中邻域窗口的索引.

2.2 多相关性 Laplacian 正则化项构造方法

假设任一像素的混合参数 α 都可以通过其邻域像素的 α 值的线性加权组合来预测. 即有如下局部线性模型:

$$\forall i \in N_m, o_{im} = \sum_{j \in N_m} w_{ij}^{(m)} \alpha_j^{(m)} \quad (5)$$

其矩阵形式为 $o_{im} = (\mathbf{w}_{-i}^{(m)})^T \boldsymbol{\alpha}_{-i}^{(m)}$, 其中,

$$(\mathbf{w}_{-i}^{(m)})^T = [w_{i,1}^{(m)}, \dots, w_{i,i-1}^{(m)}, w_{i,i+1}^{(m)}, \dots, w_{i,|N_m|}^{(m)}], (\boldsymbol{\alpha}_{-i}^{(m)})^T = [\alpha_1^{(m)}, \dots, \alpha_{i-1}^{(m)}, \alpha_{i+1}^{(m)}, \dots, \alpha_{|N_m|}^{(m)}].$$

而模型参数 $w_{ij}^{(m)}$ 通过解如下优化问题获得:

$$\min_{w_{ij}^{(m)}} \left\| \mathbf{C}_i - \sum_{j \in N_m} w_{ij}^{(m)} \mathbf{C}_j \right\|^2 \quad (6)$$

其中, $w_{ij}^{(m)}$ 满足约束条件 $\sum_{j \in N_m} w_{ij}^{(m)} = 1, \forall j \notin N_m, w_{ij}^{(m)} = 0$.

为求解公式(6),对每个像素点 \mathbf{C}_i 引入局部 Gram 矩阵 $\mathbf{M}_i^{(m)}, \mathbf{M}_i^{(m)} = (\mathbf{C}_i \mathbf{1}^T - \mathbf{C}_i^{(m)})^T (\mathbf{C}_i \mathbf{1}^T - \mathbf{C}_i^{(m)})$, 其中, $\mathbf{1}$ 是 $|N_m|$ 维的全 1 向量, $\mathbf{C}_i^{(m)}$ 是 $d \times |N_m|$ 矩阵, 每列由邻域 N_m 中的像素点构成, 对于彩色图像, $d=3$. 因此, 公式(6)中的 $w_{ij}^{(m)}$ 有如下闭解形式:

$$\mathbf{w}_{-i}^{(m)} = \frac{(\mathbf{M}_i^{(m)})^{-1} \mathbf{1}}{\mathbf{1}^T (\mathbf{M}_i^{(m)})^{-1} \mathbf{1}} \quad (7)$$

一旦获得 $w_{ij}^{(m)}$, 将其代入公式(5), 进而公式(4)的具体形式如下:

$$J_s(\boldsymbol{\alpha}) = \sum_m \left\| \boldsymbol{\alpha}_m - \mathbf{o}_m \right\|^2 = \sum_m \left\| \begin{bmatrix} \alpha_1^{(m)} \\ \alpha_2^{(m)} \\ \vdots \\ \alpha_{|N_m|}^{(m)} \end{bmatrix} - \begin{bmatrix} (\mathbf{w}_{-1}^{(m)})^T \boldsymbol{\alpha}_{-1}^{(m)} \\ (\mathbf{w}_{-2}^{(m)})^T \boldsymbol{\alpha}_{-2}^{(m)} \\ \vdots \\ (\mathbf{w}_{-|N_m|}^{(m)})^T \boldsymbol{\alpha}_{-|N_m|}^{(m)} \end{bmatrix} \right\|^2 = \sum_m \left\| \boldsymbol{\alpha}_m - \mathbf{W}^{(m)} \boldsymbol{\alpha}_m \right\|^2 = \sum_m \boldsymbol{\alpha}_m^T (\mathbf{I} - \mathbf{W}^{(m)})^T (\mathbf{I} - \mathbf{W}^{(m)}) \boldsymbol{\alpha}_m \quad (8)$$

这里, \mathbf{I} 是单位矩阵, $\mathbf{W}^{(m)}$ 是 $|N_m| \times |N_m|$ 矩阵, 第 i 行由 $(\mathbf{w}_{-i}^{(m)})^T$ 构成, 对角元素为 0. 定义 $\mathbf{L}_m = (\mathbf{I} - \mathbf{W}^{(m)})^T (\mathbf{I} - \mathbf{W}^{(m)})$, 其第 (i, j) 个矩阵元素为 $\delta_{ij} - W_{ij}^{(m)} - W_{ji}^{(m)} + \sum_k W_{ki}^{(m)} W_{kj}^{(m)}$, 其中, δ_{ij} 当 $i=j$ 时为 1, 否则为 0. 因此公式(8)最终写成

$$J_s(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{L} \boldsymbol{\alpha} \quad (9)$$

其中, \mathbf{L} 是 $n \times n$ 矩阵, 由公式(8)可知, 其 i 行 j 列元素为

$$\sum_{m|(i,j) \in N_m} \left(\delta_{ij} - W_{ij}^{(m)} - W_{ji}^{(m)} + \sum_k W_{ki}^{(m)} W_{kj}^{(m)} \right) \quad (10)$$

2.3 数据拟合约束项定义及对应的景物提取求解方法

下面给出的两种数据约束项构造, 分别对应了两种求解 $\boldsymbol{\alpha}$ 的方法:

直接求解方法(LLN-d). 数据约束项的一种直接定义方式就是要求满足用户初始所提供的部分前、背景指示约束. 即对标记为前景的像素, 置其 $\alpha=1$; 对标记为背景的像素, 置其 $\alpha=0$.

结合正则化项公式(9), 则基于局部线性学习的景物提取直接解法就是求解如下能量最小化问题:

$$J(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{L} \boldsymbol{\alpha} + (\boldsymbol{\alpha} - \mathbf{y})^T \mathbf{E} (\boldsymbol{\alpha} - \mathbf{y}) \quad (11)$$

其中, $\mathbf{y} \in \mathbf{R}^n$, 对角矩阵 $\mathbf{E} \in \mathbf{R}^{n \times n}$. 对已标记为前景和背景的像素, 向量 \mathbf{y} 的对应元素取其初始 α 值, 其余元素全置为 0; 对应 \mathbf{E} 的对角元素取常数 λ , 其余对角元素全置为 0. 因此, 公式(11)的求解相当于直接解一个线性系统:

$$(\mathbf{L} + \mathbf{E}) \boldsymbol{\alpha} = \mathbf{E} \mathbf{y} \quad (12)$$

迭代求解方法(LLN-i). 数据拟合约束项的另一种定义方式为像素颜色观测值与其预测值的差.

将图像像素划分为完全前景标记部分 V_F , 完全背景标记部分 V_B 和未知标记部分 V_U , 并记 $V_K = V_F \cup V_B$. 利用文献[1,5]的抽样技术, 像素预测值由 V_F 和 V_B 中的像素样本结合图像合成基本方程(1)得到. 即, 对于任意未知像素 p , 沿着前景区域和背景区域的边界分别选出 N 个与像素 p 最接近(按颜色值的欧氏度量)的前、背景像素. 每一对前、背景样本构成了像素 p 颜色值的估计, 因此, 数据拟合项可以如下定义:

$$J_d = \sum_{p \in V_U} \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|C_p - \alpha_p F_p^i - (1 - \alpha_p) B_p^j\|^2 / \sigma_p^2 \right) \quad (13)$$

这里, F_p^i 是指 V_F 中的前景样本颜色值, B_p^j 是指 V_B 中的背景样本颜色值, C_p 是像素 p 的实际颜色值. σ_p^2 是 C_p 与 $C_p - \alpha_p F_p^i - (1 - \alpha_p) B_p^j$ ($i, j \in \{1, 2, \dots, N\}$) 距离的方差. 本文取 $N=5$, α_p 初值取为 0.5.

将公式(9)中的 L 按已知标记像素部分 V_K 和未知标记像素部分 V_U 进行对应分解, 则公式(9)写成

$$J_s(\alpha_U) = \begin{bmatrix} \alpha_K^T & \alpha_U^T \\ \mathbf{R}^T & \mathbf{L}_U \end{bmatrix} \begin{bmatrix} \mathbf{L}_K & \mathbf{R} \\ \mathbf{R}^T & \mathbf{L}_U \end{bmatrix} \begin{bmatrix} \alpha_K \\ \alpha_U \end{bmatrix} = \alpha_K^T \mathbf{L}_K \alpha_K + 2\alpha_U^T \mathbf{R}^T \alpha_K + \alpha_U^T \mathbf{L}_U \alpha_U \quad (14)$$

其中, α_U 表示需要我们求解的未知标记像素 α 值构成的向量, α_K 是一个由 1 和 0 构成的二值向量, 对应 V_F 和 V_B 中像素的 α 值. 实际上, α_K 给出了求解 α 需要满足的边界条件.

令 $x_p^i = C_p - F_p^i$, $y_p^j = C_p - B_p^j$, $i, j=1, 2, \dots, N, p=1, 2, \dots, |V_U|$. 将公式(13)写成矩阵形式:

$$J_d(\alpha_U) = \frac{1}{N} \alpha_U^T \mathbf{P} \alpha_U + \frac{2}{N^2} \alpha_U^T \mathbf{S} (\mathbf{1} - \alpha_U) + \frac{1}{N} (\mathbf{1} - \alpha_U)^T \mathbf{T} (\mathbf{1} - \alpha_U) \quad (15)$$

其中,

$$\begin{aligned} \mathbf{P} &= \text{diag} \left(\sum_{i=1}^N x_1^i \cdot x_1^i / \sigma_1^2, \dots, \sum_{i=1}^N x_{|V_U|}^i \cdot x_{|V_U|}^i / \sigma_{|V_U|}^2 \right), \\ \mathbf{S} &= \text{diag} \left(\sum_{i=1}^N \sum_{j=1}^N x_1^i \cdot y_1^j / \sigma_1^2, \dots, \sum_{i=1}^N \sum_{j=1}^N x_{|V_U|}^i \cdot y_{|V_U|}^j / \sigma_{|V_U|}^2 \right), \\ \mathbf{T} &= \text{diag} \left(\sum_{i=1}^N y_1^i \cdot y_1^i / \sigma_1^2, \dots, \sum_{i=1}^N y_{|V_U|}^i \cdot y_{|V_U|}^i / \sigma_{|V_U|}^2 \right), \end{aligned}$$

$\mathbf{1}$ 是全 1 向量.

由正则化项公式(14)和数据拟约束项公式(15)得到能量方程:

$$J(\alpha_U) = J_s(\alpha_U) + \lambda_2 J_d(\alpha_U) \quad (16)$$

其中, 参数 λ_2 动态设定, 以强调数据项或光滑项的作用. 本文取 $\lambda_2 = e^{(t-\theta)^3}$, 其中, t 是迭代步数, θ 设置为最初的几个迭代步应让光滑项的权重尽量大, 使得除边缘外的大部分区域尽量光滑, 其后的迭代应让数据拟合项占主导作用, 以便得到更清晰的边缘, 典型地取 θ 为 6.4.

对公式(16)关于 α_U 求导, 导致求解如下线性系统问题:

$$\left(\mathbf{L}_U + \lambda_2 \left(\frac{1}{N} \mathbf{P} + \frac{1}{N} \mathbf{T} - \frac{2}{N^2} \mathbf{S} \right) \right) \alpha_U = \lambda_2 \left(\frac{1}{N} \mathbf{T} \mathbf{1} - \frac{1}{N^2} \mathbf{S} \mathbf{1} \right) - \mathbf{R}^T \alpha_K \quad (17)$$

按上述方法构造能量公式(16), 进而由公式(17)求解未知标记部分的 α_U 完成一次迭代计算. 因此, 完整的迭代解法求解步骤为:

- a) 按上述方法构造公式(13)和公式(14), 再由公式(17)求解;
- b) 由公式(17)求得 α_U 后, 将其中 $\alpha_p \geq 0.99$ 的像素 p 划归到 V_F 中, 并标记像素 p 为前景; $\alpha_p \leq 0.01$ 的像素 p 划归到 V_B 中, 并标记像素 p 为背景;
- c) 反复执行步骤 a) 和步骤 b), 直至不再有像素被划入 V_F 和 V_B 中为止.

迭代过程实现由粗到精地逐步求取 α 值. 同时, 按照文献[4]的方法, 可以从 V_F 和 V_B 中选择一对样本作为 V_U 中的像素 q 的前、背景成分的估计, 该对前、背景样本使得如下拟合误差最小:

$$(F_q, B_q) = \arg \min_{F_q^i, B_q^j} \|C_q - \alpha_q F_q^i - (1 - \alpha_q) B_q^j\| \quad (18)$$

3 相关 Laplacian 方法在景物提取中的推广应用

直接求解方法一次性求解 α , 结果不便再次优化; 而迭代传播求解结合抽样技术逐步求精获得 α 的解. 更有

意义的是,利用上述迭代技术可以推广和改进一类只具有单相关性和重相关性的 Laplacian 应用于景物提取任务中.按照 3 种 Laplacian 构造方式,相关的景物提取技术归纳为:

1) 单相关性 Laplacian 的方法. Random walk 方法^[7]使用高斯核、Global Easy 方法^[5]利用 $1/\|C_i - C_j\|$ 来计算任何一对相邻点 (i, j) 的权值.由第 1 节可知,由于这种 Laplacian 构造方式导致相邻像素标记信息按等权重传播,无法依据当前局部几何变化进一步区分标记信息的成分比例,因此在必要的光滑化前景区域和背景区域时,却导致景物目标的边缘无法保持而模糊.

利用 Global Easy 方法中构造的光滑项 Laplacian,结合公式(12)求解,给出一个与 Global Easy 迭代方法对应的直接求解方法,记该方法为 Easy- d .反过来,对于 Random walk 方法中构造的 Laplacian 采用迭代传播求解进行推广,记该推广方法为 LL1- i .

2) 重相关性 Laplacian 的方法.构造如式 $L=(I-W)^T(I-W)$ 的重相关性 Laplacian 已成功应用于图像分割^[14],但对景物提取的应用未见报道.这里,我们将其按照直接解法和迭代解法推广应用于景物提取,推广方法分别记为 LL2- d 和 LL2- i .由第 1 节可知,任何一对相邻点的相似性权值计算,依据它们各自为中心的局部邻域结构的不同而不同.直觉上,该类方法的景物提取效果必定优于单相关性 Laplacian 的方法.

3) 多相关性 Laplacian 的方法. Closed-form^[8], 本文的 LLN- d 和 LLN- i 均属于此类方法.由第 1 节可知,该类方法在构造 Laplacian 时,任何一对相邻点的相似性权值计算会随着周围任何局部邻域(只要它们落入该邻域)几何结构的变化而变化,因此,它们能够约束标记成分以更精细的方式传播,在光滑前景和背景区域时能够很好地保持边缘等细节.当然,同样可以将文献[8]中构造的 Laplacian 结合迭代方法来求解(记为 Closed- i).

上述各种景物提取的 Laplacian 方法按其构造方式、求解方法和所包含的邻点对相似性权值计算次数进行分类,见表 1.每列的方法都是采用同一个 Laplacian 矩阵 L ,区别只是求解方法不同而已.

Table 1 Classification of matting methods based on Laplacian

表 1 各种景物提取 Laplacian 方法分类汇总表

	Laplacian classification				
	Single-Correlation Laplacian		Double-Correlation Laplacian	Multi-Correlation Laplacian	
Direct solver	Easy- d	Random walk ^[7]	LL2- d	Closed-Form ^[8]	LLN- d
Iterative solver	Global Easy ^[5]	LL1- i	LL2- i	Closed- i	LLN- i
#weights for each pair of adjacent points	1		2	>2	

4 结果分析与比较

所提方法及与相关方法在自然图像中提取景物目标的效果比较,都是基于用户提供的同一个稀疏前、背景条带约束下获得的.

首先,给出各种典型方法提取前景目标的例子.由图 2 可见,本文的迭代解法(如图 2(j)所示)在这个例子上获得了最好的结果. Bayesian 方法由于缺乏先验似然 $P(\alpha)$ 的光滑性约束(文献[2]省略该项)导致大块区域不能光滑. Poisson 方法利用图像灰度梯度场作为 α 梯度的估计,导致不具有边缘保持的光滑扩散,同时又缺乏数据拟合约束项,因此, Poisson 方法只有在用户提供精细的 trimap 划分下方能奏效. Geodesic matting 方法(如图 2(f)所示)容易留下前后两次前、背景分布估计的痕迹. Global Easy 方法与 Random walk 方法一样也是构造单相关性 Laplacian,因采用迭代优化求解技术而获得了更好的结果(如图 2(c)所示).

其次,同类方法比较以及推广单相关性 Laplacian 和重相关性 Laplacian 方法求解 α 的实验验证. 考察表 1 中的景物提取方法在图 3(a)中提取小狗的例子. 对于单相关性 Laplacian 方法(如图 3(b)~图 3(e)所示), Easy- d 与 Random walk 均无法获得理想的结果. 而相应的 Global Easy 与 LL1- i 方法由于选择迭代求解,使得结果获得极大改善. 重相关性 Laplacian 方法(如图 3(f)、图 3(g)所示)显然比单相关性 Laplacian 方法的景物提取结果更优. 虽然 LL2- d 在边缘细节处存在一些模糊,但迭代解法 LL2- i 的结果是令人满意的. 对于多相关性 Laplacian 方法,由于其 Laplacian 正则化项构造充分考虑了局部邻域几何结构的变化,因此,本文方法与文献[8]的闭解方法一样,均能获得理想的景物提取结果(如图 3(h)~图 3(k)所示). 尽管 Robustmatting 将景物提取问题转化为一个随机

步行优化图标记问题,但使用与文献[8]完全相同的 Laplacian,因此从信息传播角度,这个优化策略与闭解方法没有本质上的差异,结果如图 3(l)所示.

值得注意的是,3 类 Laplacian 的一次迭代求解的结果如图 3(m)~图 3(o)所示,相当于对应的直接求解结果(Global Easy 一次迭代的结果与图 3(b)相似).换句话说,迭代解法可以看成是对直接解法结果的不求精过程.

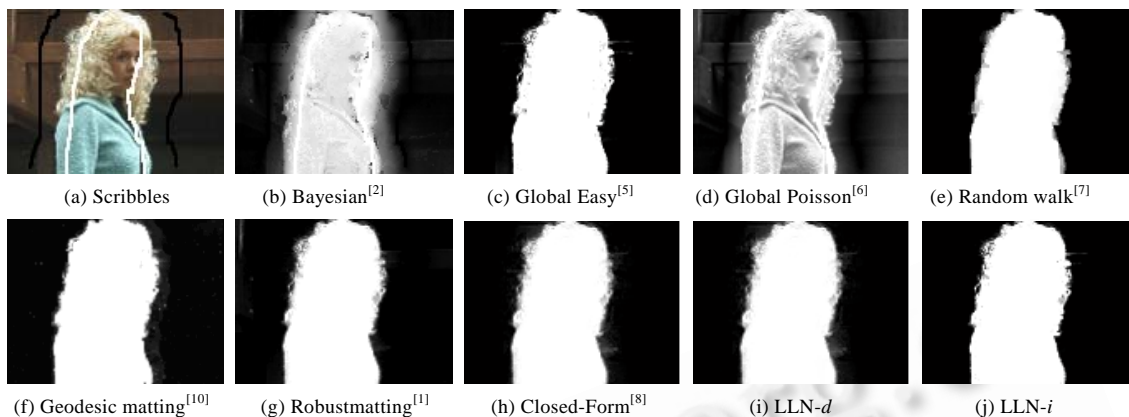


Fig.2 Extracted α mattes by some representative methods
图 2 几个典型方法获得的景物提取混合参数 α 的例子

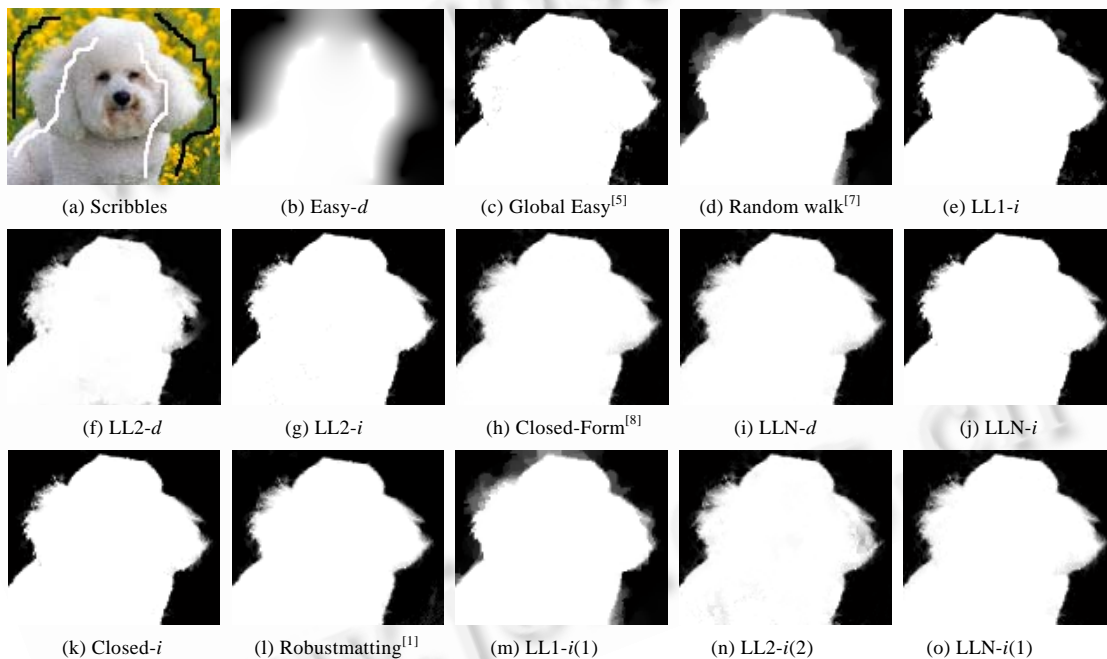


Fig.3 Extracted α mattes by three types of Laplacian with direct and iterative solvers
图 3 3 种 Laplacian 对应的直接和迭代求解 α 的例子

再借助文献[1]提供的景物提取基准测试图,如图 4 所示,从左至右编号为 T1~T8(<http://juew.org/data/data.htm>),进一步给出上述同类方法的误差平方和(SSE)性能比较结果.对于大图像,我们利用文献[8]的计算方法:通过下采样图像来减小线性系统公式(12)和公式(17)的规模,求解一个低分辨率图像的 α 值,然后通过内插由粗到精地恢复原始图像 α 值.8 幅图像及每幅图像由上述除 Easy-d 外的 10 种方法获得 α 值后计算 SSE(sum of squares errors),如图 5 所示,同时,图 6 给出了 T5 的结果图像.



Fig.4 Test images with scribbles and their α ground-truth mattes

图 4 测试图像及对应的 α 基准图像

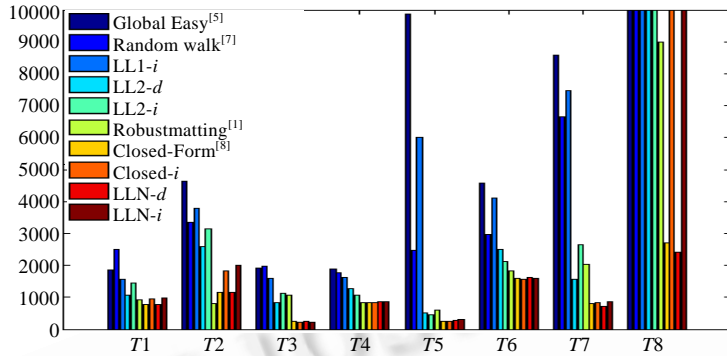


Fig.5 A SSE comparison of α mattes produced by the same type of matting methods

图 5 同类方法求解 α 的 SSE 结果比较

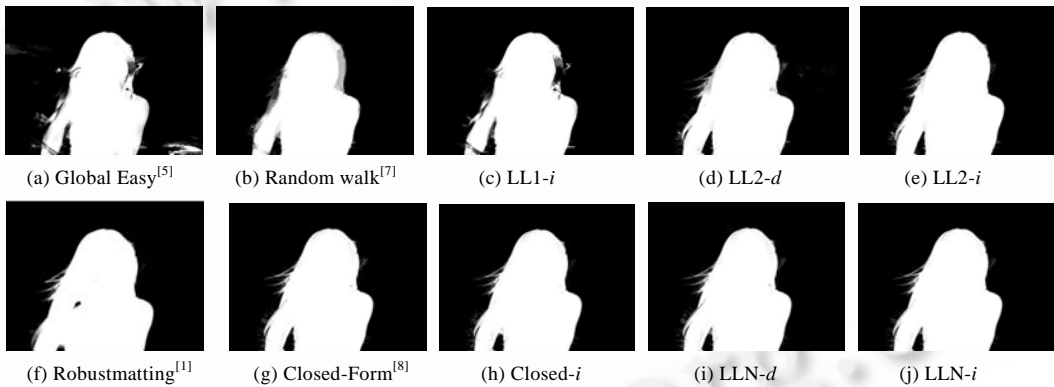


Fig.6 α mattes extracted by the algorithms with different Laplacians for T5

图 6 各种 Laplacian 方法求解 T5 的 α 结果

可见,多相关性 Laplacian 的方法几乎在所有图像上均获得满意的结果.Closed-form 和 LLN-d 在 T1~T6 这 6 幅图像结果相当,在 T7 和 T8 上 LLN-d 有更好的表现,并且均优于 Robustmatting.意味着 LLN-d 在细节丰富、前、背景易混淆的图像中适应局部结构变化的能力具有相对优势.Robustmatting 在提供精细的 Trimap 划分的情况下能够获得精确的景物提取结果^[1],但当前景和背景部分标记稀疏时,由于很难通过抽样技术获得理想的初始 α 估计,导致提取结果的准确度下降.迭代求精方法可以进一步获得更清晰的边缘细节,但同时如图 6(h)~图 6(j)显示的那样,容易使细节变得更尖锐,所提取的景物目标倾向于硬分割的结果.虽然迭代解法 Closed-i 和 LLN-i 在 T3 和 T6 上相对于直接解法 Closed-form 和 LLN-d 的 SSE 指标均有所下降,但对于其他细节变化丰富的提取目标,迭代解法使细节尖锐化导致 SSE 性能得不到改善,甚至在 T8 上完全失败.

在 T1~T7 这 7 幅图像上,LL2-d 和 LL2-i 均能获得较好的目标提取结果,迭代解法相对于直接解法在 T4~T6

上有较大改善.LL1- i 显著优于 Global Easy,并在 $T1, T3$ 和 $T4$ 上极大地改善了其对应直接解法 Random walk 的结果.然而由于单相关性 Laplacian 的低保细节能力,导致了它们在 $T5, T7$ 和 $T8$ 上的表现很不理想.例如,从图 6(a)-图 6(c)中 $T5$ 的结果可知,Global Easy 和 LL1- i 在迭代过程中,一方面没有光滑大块背景区域,另一方面也丧失了许多细节.这些问题可以通过增加指示条带的密度和利用文献[5]的局部多次交互来对结果求精的办法加以克服.多次交互的问题超出了本文的讨论范围.

最后,由上述各种方法获得 α 后,可以通过公式(18)估计前景成分 F 和背景成分 B ,利用图像合成方程(1)就可以将上述获得的景物目标与任意背景图像合成得到新图像,图 7 给出了本文方法及一些推广方法的合成结果.



Fig.7 Some composition examples

图 7 合成新图像的例子

5 结论

本文在半监督学习框架下,依据任意一对相邻点出现在不同的邻域中时是否重新计算它们的相似性权值,将正则化 Laplacian 的构造划分为 3 种类型,并从理论上证明了多次计算不同相似性权值的 Laplacian 构造方式能够更“忠实”和准确地反映数据点的局部几何结构.此外,根据这一理论,本文还提出了一种新的基于局部学习的 Laplacian 构造方法,同时推广了一大类一次和两次计算相似性权值的 Laplacian 应用于景物提取任务中.理论分析和实验结果获得了一致的结论,即要获得理想的景物目标,则应构造具有多次相似性权值计算的 Laplacian.本文所述的这类方法主要着眼于图像局部几何的刻画而忽视了图像颜色值的统计分布,尽管文中的迭代传播求解方法利用了统计抽样技术,但在大多数情形下,对于包含多次权值计算的 Laplacian 是不必要的.将来的工作主要是如何将本文的局部几何方法与全局的统计技术进行更有效的结合,以及其在其他图像编辑任务中的推广应用.

References:

- [1] Wang J, Cohen MF. Optimized color sampling for robust matting. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Minneapolis, 2007. <http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=4270031>
- [2] Chuang YY, Curless B, Salesin DH, Szeliski R. A Bayesian approach to digital matting. In: Jacobs A, Baldwin T, eds. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2001. 264–271.
- [3] Lin SY, Pan RF, Du H, Shi JY. A survey on digital matting. Journal of Computer-Aided Design & Computer Graphics, 2007,19(4):473–479 (in Chinese with English abstract).
- [4] Wang J, Cohen MF. An iterative optimization approach for unified image segmentation and matting. In: Ma SD, Shum HY, eds. Proc. of the IEEE Int'l Conf. on Computer Vision. New York: IEEE Computer Society Press, 2005. 936–943.
- [5] Guan Y, Chen W, Liang X, Ding ZG, Peng QS. Easy matting—A stroke based approach for continuous image matting. Computer Graphics Forum, Eurographics, 2006,25(3):567–576.
- [6] Sun J, Jia JY, Tang CK, Shum HY. Poisson matting. ACM Trans. on Graphics, 2004,23(3):315–321.
- [7] Grady L, Schiwietz T, Aharon S, Westermann R. Random walks for interactive alpha-matting. In: Villanueva JJ, ed. Proc. of the 5th IASTED Int'l Conf. on Visualization, Imaging and Image Processing. Benidorm: ACTA Press, 2005. 423–429.
- [8] Levin A, Lischinski D, Weiss Y. A closed form solution to natural image matting. In: Fitzgibbon A, Taylor C, LeCun Y, eds. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2006. 61–68.

- [9] Levin A, Alex RA, Lischinski D. Spectral matting. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Minneapolis, 2007. <http://people.csail.mit.edu/alevin/papers/spectral-matting-levin-et-al-cvpr07.pdf>
- [10] Bai X, Sapiro G. A geodesic framework for fast interactive image and video segmentation and matting. In: Proc. of the 11th IEEE Int'l Conf. on Computer Vision. Rio de Janeiro, 2007. <http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=4408931>
- [11] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323-2326.
- [12] Zhu XJ, Ghaharmani Z, Lafferty J. Semi-Supervised learning using Gaussian fields and harmonic functions. In: Fawcett T, Mishra N, eds. Proc. of the 20th Int'l Conf. on Machine Learning. Menlo Park: AAAI Press, 2003. 912-919.
- [13] Zhou DY, Bousquet O, Lal TN, Weston J, Schölkopf B. Learning with local and global consistency. In: Thrum S, Saul L, Schölkopf B, eds. Advances in Neural Information Processing System 16. Cambridge: MIT Press, 2004. 321-328.
- [14] Wang F, Wang JD, Zhang CS, Shen HC. Semi-Supervised classification using linear neighborhood propagation. In: Werner B, ed. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2006. 160-167.
- [15] Wang F, Zhang CS. Label propagation through liner neighborhoods. In: Cohen WW, Moore A, eds. Proc. of the 23rd Int'l Conf. on Machine Learning. New York: ACM Press, 2006. 985-992.
- [16] Wu MR, Schölkopf B. Transductive classification via local learning regularization. In: Meila M, Shen X, eds. Proc. of the 11th Int'l Conf. on Artificial Intelligence and Statistics. Cambridge: MIT Press, 2007. 624-631.

附中文参考文献:

- [3] 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.



彭宏京(1965—),男,江西遂川人,博士,副教授,主要研究领域为模式识别,计算机视觉,数据挖掘.



张道强(1978—),男,博士,教授,主要研究领域为机器学习,模式识别,数据挖掘.



陈松灿(1962—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别,智能计算.