

全局显著结构主导下的知觉编组算法

邹 琪 罗四维 钟晶晶

(北京交通大学计算机与信息技术学院 北京 100044)

摘 要 提出一种建立在可靠的全局线索基础上的编组算法. 编组线索为反映全局显著结构的拓扑特征闭合性和平行性以及局部规律邻接性和连续性. 依据概率推理选择最显著的边缘作为种子, 依据全局依赖性选择最有可能与种子属于同一编组的边缘. 编组的形成中融入注意机制, 一方面缩小寻优空间另一方面确定各编组被检测的顺序. 在 Berkley 图像库上的实验表明, 该算法至少具有与 Ncut 和 mini-cut 相当的准确率, 特别对纹理少的图像能够有效地降低错编率与漏编率. 同时由于对边缘进行编组降低了输入数据的维数, 因此比 Ncut 和 mini-cut 更少地受到图像尺寸的限制.

关键词 知觉组织; 拓扑特征; 注意机制

中图法分类号 TP391

A Perceptual Grouping Algorithm Based on Global Salient Structure

ZOU Qi LUO Si-Wei ZHONG Jing-Jing

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044)

Abstract A grouping algorithm based on global salient structure is proposed. Grouping cues are topological properties namely parallelism and closure and local principles namely proximity and continuity. The most salient edge according to probability reference is selected as grouping seed. Edges determined by global statistical dependency are selected as subsequential ones with the most probability of being in the same group with the seed. In perceptual grouping process, attention is employed in grouping to both reduce optimal space and decide pop-out sequence of groups according to their salience. Compared with algorithms adopting local salient relations, above algorithm provides more reliable cues for nature images. This group-based attention makes the effect close to human perception. Experiments on Berkley image database show above algorithm achieves accuracy competitive to Ncut and mini-cut algorithms. It reaches lower error rate and missing rate especially on images with litter texture. Meanwhile, compared with graph cut methods grouping on pixels, the proposed algorithm grouping on edges reduces input dimensionality, therefore less restrictive in image size.

Keywords perceptual organization; topological property; attention

1 引 言

知觉编组(perceptual grouping)是生物视觉感

知最基本的功能之一, 即分辨哪些信息属于一个整体进而把这些信息组织成一个有明确景象意义的目标^[1], 在计算机视觉中是目标背景分割和检测识别的先决条件. 知觉编组在视觉中起着承上启下的作

用,输入为特征提取阶段提取的点、线段、区域块等初级视觉元素,输出为感知的目标,为高级阶段的处理如检测和识别提供对象. 编组的结果可以直接作为分割的结果,但编组是比分割更复杂的过程. 分割的结果不一定与感知目标一一对应,可能产生过分割 (over-segmentation) 或者不充分分割 (under-segmentation). 前者使完整的目标拆分成若干独立的成分,后者将不同的目标在分界线不明显的地方混淆为同一目标. 而编组考虑了知觉组织原则,能得到更有景象意义的目标.

知觉编组的概念源于认知心理学的研究,通常以格式塔心理学提出的编组原则作为线索 (cues). 传统的编组方法过于依赖最终的优化决策阶段,而忽视了前期的编组依据提取阶段所提取的线索的可靠性对编组质量的影响. 最常用的编组算法是基于图分割的方法^[2-4]. 这一类方法将图像基元间的关系表示为以基元为顶点,基元间的关联程度 (即线索) 为边的图. 最大化结合或聚类概率得到的生成图对应最终的编组. 生成图中的每个分割子图对应一个独立目标. 用基于图分割的方法实现编组面临着两个难题. 首先,图分割方法一般情况下都属于 NP 难问题,只有在特定的约束条件下才能在有限的计算量内求解. 其次,采用图表示结构使得局部关系,尤其是二元关系 (一对基元之间的关系) 更直观,更容易计算. 然而,视觉感知中全局关系是可靠性更高、更为显著的关系. 认知心理学认为,局部信息在知觉上是模糊的,只有这些局部信息作为一个有序的整体体现出全局结构时才使目标与背景有明确的景象意义^[1]. 为解决第一个难题,研究者们采用各种优化方式以使算法能够快速准确地收敛到全局最优解,如模拟退火算法^[5]、谱聚类算法^[2,4]等. 为解决第二个难题,研究者们倾向于用基于概率论的决策方式来选择二元关系

中相对比较可靠的组合,如:贝叶斯网、极大似然估计等. 文献[6]通过对形状模板的学习使二元关系的组合具有很强的鲁棒性. 文献[7]将多尺度分析的思想引入到编组中,以克服二元关系的图结构对噪声的敏感性. 这些方法本质上都是利用局部关系来发现显著结构以确定编组. 近年来,也有研究者尝试通过提取全局拓扑特征实现编组. Williamms 将随机完全场 (stochastic completion field)^[8-9] 的概念引入到拓扑特征的提取中,以同圆性和邻接性为线索形成编组,在某一点的场强是图像所有其它边缘对该点的影响的叠加,实际上是将全局关系表示为所有局部关系的和. Elder^[10]利用假设存在的先验知识,将简单性、闭合性、完整性的全局约束与局部线索结合起来,并应用在卫星遥感图像的轮廓提取上,但适用范围有限,仅限于对编组目标的形状事先了解的情况. Mahamud^[11]设计了马尔可夫过程的状态转移矩阵表示构成闭合轮廓的边缘间的依赖关系,能在遮挡、噪声的干扰下分割出完整的编组,但对目标轮廓的约束较强,只适合发现连接足够光滑的凸闭合轮廓.

鉴于现有的方法缺乏对全局关系的有效定义,而是将编组建立在可靠性低的局部线索上,本文提出新的全局显著结构的定义方式. 这里的全局关系充分考虑了闭合轮廓可能出现的各种情况,解除了对闭合轮廓形状的限制,并且这里的全局关系不是局部关系的简单线性叠加,而是根据全局显著性定义严格推导出来的真正意义上的全局关系. 正是这种全局显著结构使得该算法的编组质量高于依赖局部关系的编组算法和以其它方式定义全局关系的编组算法. 同时,受生物视觉机理的启发,提出一种基于注意机制的编组模型,缩小了寻优空间,避免了全局范围的贪心搜索. 注意机制还按照各编组的显著程度确定其被检测的顺序. 算法流程如图 1 所示.

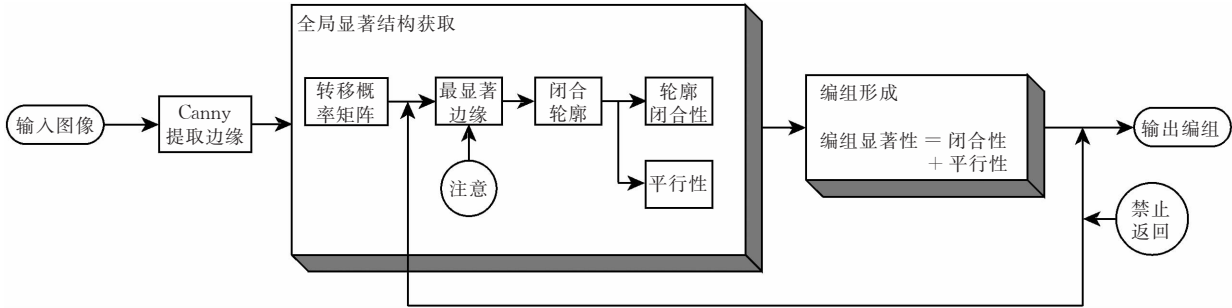


图 1 算法流程图

本文第 2 节介绍生物视觉的相关概念及全局关系的获取与表示;第 3 节描述具体的全局显著结构

主导下的编组算法 (以下简称 GSSD 算法) 以及此过程中注意机制的工作原理;实验部分对不同类别的

自然图像验证该算法,与其它算法作定量的比较,并对结果作理论分析;未来工作的方向将在总结部分讨论。

2 全局显著结构的获取与表示

格式塔心理学研究表明,人能够在不需要相关知识觉学习的情况下,自主地利用某些既定的规律组织视觉信息,从初级阶段提取的线段位置、朝向、频率等局部特征中感知到整体有序的结构^[1]. 这些结构反映了局部特征之间的拓扑关系,符合知觉编组规律. 按照这些规律形成的图形比相对无序缺少清晰规律的背景更为显著,因此从背景中突显出来并进一步分割成若干潜在的目标。

编组可以仅依靠图像边缘信息,也有利用区域特征如形状、纹理等进行编组的^[12]. 常用的编组规律包括邻接性(proximity)、平行性(parallelism)、共线性(colinearity)、良好连续性(good continuation)和闭合性(closure)等. 在图结构中这些规律被表示为一对线段间的局部关系. 比如:用一对线段的顶点间的距离衡量邻接性,斜率差异衡量共线性. 然而就整幅图像的范围看,线段对(edge-pair)之间的二元关系不足以提供充分的依据以反映哪些线段属于同一个编组,一个组的总体关系也不等于组内所有线段对之间的二元关系的总和. 如图 2 所示,线段 e_1, e_2, e_3 属于一个编组, e_3 和 e_4 属于另一个编组. 如果利用二元关系来判断编组, e_1, e_3 的共线性和邻接性对 e_1, e_3 属于同一编组的支持程度比 e_1, e_2 的连续性对 e_1, e_2 属于同一编组的支持程度高,这样就会误判 e_1, e_3 属于同一编组. 但是利用全局关系,这种误判出现的几率就小得多. 如 e_1, e_2, e_5 (满足连续性)属于同一编组的可能性远大于 e_1, e_3, e_5 (不满足任何编组规律). 同时, e_4, e_6, e_7, e_8, e_9 和 e_3 (满足闭合性)属于同一编组的可能性远大于 e_1 和 e_3 属于同一编组的可能性。

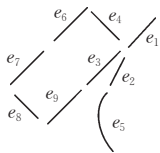


图 2 全局关系比二元关系提供更可靠的编组依据

正是由于全局关系的可靠性高,本文采用线段所反映的全局关系作为编组的线索. 线段来源于 Canny 算子提取的图像边缘. 为每条边缘保留对比

度、方向、位置和长度四类特征. 其中,对比度为边缘所在处图像灰度变化的幅度. 用到的编组规律包括由局部性质邻接性和连续性通过概率推理的方法得到的全局结构以及全局拓扑性质闭合性和平行性(这两种规律的对象为一组边缘序列形成的首尾相连的完整封闭的轮廓)。

2.1 闭合性

用边缘序列体现的全局拓扑关系来判断闭合性. 首先构造转移概率矩阵来描述闭合轮廓中边缘的连接关系,然后利用转移矩阵的特征向量选出最显著的边缘,接下来以最显著的边缘为起点提取闭合轮廓,即顺序查找最有可能与起点属于同一闭合轮廓的边缘,最后对提取的闭合轮廓计算闭合程度,作为编组尺度之一。

2.1.1 转移概率矩阵

两条边缘间的转移概率反映了这两条边缘位于同一条闭合轮廓里的可能性. 考虑闭合轮廓可能出现的 3 种情况:光滑的连接、突变的连接(如 L 形或 T 形拐角)和遮挡引起的断裂^[3]. 边缘 e_1 和 e_2 间的转移概率 $p(e_1 e_2)$ 可按照式(1)计算,即

$$p(e_1 e_2) = p(e_1 e_2 | r_1) p(r_1) + p(e_1 e_2 | r_2) p(r_2) + p(e_1 e_2 | r_3) p(r_3) \quad (1)$$

其中, r_1, r_2 和 r_3 分别代表光滑的连接、突变的连接和遮挡引起的断裂 3 种情况. 对每一种情况又可转化为求其后验概率

$$p(e_1 e_2 | r_i) = \frac{p(r_i | e_1 e_2) p(e_1 e_2)}{p(r_i)} \quad (2)$$

由于 r_1, r_2, r_3 是 e_1 和 e_2 属于同一闭合轮廓的必要但非充分条件(不属于同一条闭合轮廓的边缘也可能存在光滑、突变或遮挡的连接关系),故 r_i 的全概公式为

$$p(r_i) = p(r_i | e_1 e_2) p(e_1 e_2) + p(r_i | \overline{e_1 e_2}) p(\overline{e_1 e_2}) \quad (3)$$

其中, $\overline{e_1 e_2}$ 表示边缘 e_1 和 e_2 不属于同一条闭合轮廓. 将式(3)代入式(2),可得

$$p(e_1 e_2 | r_i) = \frac{1}{1 + \frac{p(r_i | \overline{e_1 e_2}) p(\overline{e_1 e_2})}{p(r_i | e_1 e_2) p(e_1 e_2)}} \quad (4)$$

这里取 $p(\overline{e_1 e_2})/p(e_1 e_2)$ 为 5, 因为对每条边缘 e_1 , 只考虑与其距离最近的 6 条边缘之间的关系. 按照平均地估计,在这 6 对边缘中,跟 e_1 属于同一闭合轮廓的边缘数目与跟 e_1 不属于同一闭合轮廓的边缘数目之比为 1 : 5. 此外,用邻接性和连续性作为衡量边缘间连接关系 r_i 的尺度,则

$$p(r_i | e_1 e_2) = p(r_i^{\text{prox}} | e_1 e_2) + p(r_i^{\text{cont}} | e_1 e_2) \quad (5)$$

邻接性 r_i^{prox} 定义为两条边缘最靠近的端点间的距离与较短的边缘长度之间的比值, 连续性 r_i^{cont} 定义为边缘方向角间的差异. 邻接性和连续性的条件概率均用高斯函数模拟, 只是各自的方差 σ_i^o (其中 $o = \text{prox}$ 或 cont) 不同:

$$p(r_i^o | e_1 e_2) = \frac{1}{\sqrt{2\pi}\sigma_i^o} e^{-\frac{(r_i^o)^2}{2(\sigma_i^o)^2}} \quad (6)$$

$i = 1, 2, 3, o = \text{prox}, \text{cont}$

在边缘 e_1 和 e_2 处于同一轮廓的前提下, 若边缘间是光滑的连接, 则距离和方向差都应较小, 此时 $\sigma_1^{\text{prox}} = 1/8, \sigma_1^{\text{cont}} = \pi/18$; 若是突变的连接, 则距离应较小而方向差较大, 此时 $\sigma_2^{\text{prox}} = 1/8, \sigma_2^{\text{cont}} = \pi/2$; 若是遮挡引起的断裂, 则距离应较大而方向差较小, 此时 $\sigma_3^{\text{prox}} = 1/2, \sigma_3^{\text{cont}} = \pi/18$. 当边缘 e_1 和 e_2 不属于同一轮廓时, 仍然用邻接性和连续性作为衡量边缘间连接关系 r_i 的尺度

$$p(r_i | \overline{e_1 e_2}) = p(r_i^{\text{prox}} | \overline{e_1 e_2}) + p(r_i^{\text{cont}} | \overline{e_1 e_2}) \quad (7)$$

由于此时边缘间的连接关系是随机的, 距离和方向差都近似为均匀分布, 故式(7)中 $p(r_i^{\text{cont}} | \overline{e_1 e_2}) = 1/\pi$, 即不属于同一轮廓时, 方向差为 $[0, \pi)$ 之间的等概率分布; $p(r_i^{\text{prox}} | \overline{e_1 e_2}) = r_i^{\text{prox}}/\sqrt{2}R$ 即距离与图像中边缘分布的密集度有关, 对一幅 $M \times N$ 的图像 $R = \max\{M, N\}$.

以上是转移概率的计算方法. 转移概率矩阵 \mathbf{P} 是以 $p_{ij} = p(e_i, e_j)$ 为元素的方阵. 由于只需考虑某一条边缘与其最近邻的 6 条边缘之间的转移概率, 因此 \mathbf{P} 是稀疏矩阵, 并且是对角元素为 0 的对称矩阵. 与大多数其它编组方法一样, 转移概率是针对一对边缘间的距离、方向差等局部关系, 只是我们的计算方法更充分地考虑了边缘间可能出现的情况. 然而, 大多数方法^[3,5,10] 对于闭合性的计算进行到这一步就结束了, 接下来是编组的决策阶段, 即找到一条闭合回路, 使得这条路径上的所有相邻边缘间的转移概率之和为最大, 满足这样条件的回路就对应着最显著的闭合轮廓. 如引言部分所述, 这种寻优方法计算量很大, 可能是 NP-hard 问题, 并且这样得到的闭合性实质是局部关系之和. 与上述方法不同的是, 我们依据这些局部关系计算出全局显著结构来产生闭合轮廓. 同时, 模拟注意机制的工作原理, 首先确定全局显著性最高的边缘, 然后以这条边

缘为起点在其邻域内作局部搜索, 找到闭合程度最高的边缘序列, 这样与从任意起点开始作贪心搜索的图分割方法相比, 极大降低了计算量.

2.1.2 最显著边缘

定义边缘的全局显著性为闭合轮廓经过这条边缘的概率, 那么最显著的边缘就是经过它的所有闭合轮廓的概率之和最大的边缘

$$E = \{e_i | i = \arg \max_i \sum_{j=3}^{\text{tol}} p(e_i \in CC^j)\} \quad (8)$$

记 E 为最显著的边缘, CC^j 为长度为 j (即由 j 条边缘组成) 的闭合轮廓, tol 为图像中的边缘总数. 如果把闭合轮廓看作是可以循环重复的边缘序列, 那么经过边缘 e_i 的所有闭合轮廓可看作组成了一条无限长的闭合轮廓^[11], 所以经过边缘 e_i 的任意长度的闭合轮廓的概率之和就等于经过边缘 e_i 的无限长的闭合轮廓的概率

$$E = \{e_i | i = \arg \max_i \lim_{j \rightarrow \infty} p(e_i \in CC^j)\} \quad (9)$$

为计算经过 e_i 无限长的闭合轮廓的概率, 首先看一个引理: 转移矩阵的 n 次幂 \mathbf{P}^n 的第 i 个对角元素 p_{ii}^n 表示经过边缘 e_i 的长度为 n 的闭合轮廓的概率 (证明见附录). 因此经过 e_i 无限长的闭合轮廓的概率可表示为

$$E = \{e_i | i = \arg \max_i \lim_{n \rightarrow \infty} p_{ii}^n\} \quad (10)$$

根据矩阵分析里的定理^[13] 对于正实矩阵有

$$\lim_{n \rightarrow \infty} (\mathbf{P}/\lambda)^n = \mathbf{x} \cdot \tilde{\mathbf{x}} \quad (11)$$

其中 $\mathbf{x}, \tilde{\mathbf{x}}$ 分别为矩阵 \mathbf{P} 的最大特征值 λ 对应的右特征向量和左特征向量. 对于对称矩阵 \mathbf{P} , 有 $\mathbf{x}^T = \tilde{\mathbf{x}}$. 由于只需比较边缘显著性的相对大小, 式(10)除以常数 λ 不影响对最显著边缘的选取

$$E = \{e_i | i = \arg \max_i \lim_{n \rightarrow \infty} (\mathbf{P}/\lambda)^n_{ii}\} = \{e_i | i = \arg \max_i (x_i \cdot \tilde{x}_i)\} = \{e_i | i = \arg \max_i (x_i^2)\} \quad (12)$$

x_i, \tilde{x}_i 分别为 $\mathbf{x}, \tilde{\mathbf{x}}$ 的第 i 个元素. 这样, 转移矩阵的最大特征值对应的特征向量中的最大元素就标记着最显著的边缘. 选取全局最显著的边缘作为闭合轮廓的种子, 这一步骤与生物视觉中注意机制的功能相似. 这样, 模拟注意机制将初始状态限定在最有可能形成闭合轮廓的边缘中.

2.1.3 闭合轮廓

闭合轮廓的概率为闭合轮廓上所有相邻边缘的转移概率的乘积. 因此, 经过边缘 e_j 与 e_i 且长度为 n 的闭合轮廓 CC_{ij}^n 的概率为 $p_{ji}^{n-k} p_{ij}^k$ ($k = 1, 2, \dots, n-1$), 即从 e_j 出发经过 $n-k$ 步到达 e_i 的概率 p_{ji}^{n-k} , 与

再从 e_i 出发经过 k 步返回 e_j 的概率 p_{ij}^k 的乘积. 与种子边缘 e_i 最有可能位于同一闭合轮廓的边缘是

$$\begin{aligned} \text{neighb}(e_i) &= \{e_b \mid b = \arg \max_b \lim_{n \rightarrow \infty} p(e_b \in CC_{ib}^n)\} \\ &= \{e_b \mid b = \arg \max_b \lim_{n \rightarrow \infty} p_{bi}^{n-k} p_{ib}^k\} \quad (13) \end{aligned}$$

同样根据“除以常数因子 λ 不影响对最大值的选择”这一原则以及矩阵分析中用到的定理有

$$\begin{aligned} \text{neighb}(e_i) &= \{e_b \mid b = \arg \max_b \lim_{n \rightarrow \infty} (\mathbf{P}/\lambda)_{bi}^{n-k} (\mathbf{P}/\lambda)_{ib}^k\} \\ &= \{e_b \mid b = \arg \max_b (x_b \tilde{x}_i) (\mathbf{P}/\lambda)_{ib}^k\} \\ &= \{e_b \mid b = \arg \max_b (x_b \tilde{x}_i) p_{ib}^k\} \quad (14) \end{aligned}$$

其中, x_b, \tilde{x}_i 分别为矩阵 \mathbf{P} 的最大特征值 λ 对应的右特征向量的第 b 个元素和左特征向量的第 i 个元素. 实际运算中取 $k=1$, 即只考虑闭合轮廓连续经过 e_i 和 e_b (中间没有其它边缘) 的情形, 也就是以种子 e_i 为起点, 找到与之最有可能位于同一闭合轮廓并且与之紧相邻的边缘, 作为下一边缘. 每一次循环均按式(14)寻找当前边缘的下一边缘, 如此下去, 直到又找回到起点, 闭合轮廓便形成了. 因为 \mathbf{P} 为对称矩阵, 可能出现以下情况: 当前边缘 e_i 的下一边缘为 e_r , 而 e_r 的下一边缘又为 e_i . 出现这种自循环的情况时, 将式(14)调整为寻找使 $(x_b \tilde{x}_i) p_{ib}^k$ 第二大的参数 b , 即与当前边缘第二有可能位于同一闭合轮廓并且与之紧相邻的边缘, 作为下一边缘.

闭合轮廓形成后, 计算由闭合性引起的显著程度为

$$SC = \left(\prod_{e_i, e_j \in CC} p_{ij} \right)^{1/u}, \quad e_j = \text{neighb}(e_i) \quad (15)$$

其中, u 为闭合轮廓 CC 的长度. 开 u 次方是为避免长度越大的轮廓显著性越低 (因为长度越大的轮廓相乘的 p_{ij} 个数越多, 而 $0 < p_{ij} < 1$).

2.2 平行性

定义平行性^[12]为

$$SP = \frac{\sum_{l_i \in L_s} \|l_i\|}{\sum_{l_j \in L} \|l_j\|} \quad (16)$$

其中, L 为构成某一闭合轮廓的所有边缘线段的集合; L_s 表示属于 L , 并且与 L 中至少一条线段的方向差异在 5° 之内的边缘集合; $\|l_i\|$ 表示边缘线段的长度. 显然, SP 的取值范围在 0 和 1 之间. SP 为 1 时, 说明该区域是一个完全平行的结构 (如矩形或平行多边形). 大多数人造目标具有平行结构, 因此, 利用平行性容易探测到自然场景中的人造目标.

至此, 我们定义了闭合性和平行性两种全局拓

扑关系的表示, 闭合性又是由局部邻接性和连续性来度量. 从中可以看出全局关系与局部关系的区别. 局部关系描述一对基元之间的关系, 忽略了其它基元可能对这一对基元产生的影响; 全局关系可能由二元关系组成, 但体现的是群体内所有基元之间的依赖关系, 是所有基元作为一个整体表现出的共有的性质. 比如: 局部邻接性和连续性描述边缘两两间的空间几何关系, 仅以此判断两条边缘属于一组显然是不可靠的. 而闭合性描述一个边缘序列的整体属性, 能够作为分组的可靠依据.

3 编组形成

编组的显著性定义为

$$H = \alpha SC + \beta SP \quad (17)$$

其中, α, β 分别是闭合性和平行性的权值. 将两种多元关系归一化到同一值域范围后, 采用基于图像内容的全局放大法^[14]设定权值. 一个编组产生后, 根据注意机制的禁止返回 (inhibition of return) 原则^[15] (即一旦一个对象被注意过, 那么这个对象在以后的观察中将不被列为查找范围), 属于该编组的边缘均从原图中删除. 再从余下的边缘中获取全局显著结构, 按照编组算法产生新的编组. 如此循环, 不断产生新的编组, 直到余下的边缘无法再形成编组. 比较各编组的显著性, 并按显著程度的降序排列, 显著程度高的编组优先受到注意.

算法的运算量主要来自闭合性计算部分, 其时间复杂度最坏情况下为 $O(N) + O(N^3) + O(N)$ (转移概率矩阵 + 最显著边缘 + 闭合轮廓), 最优情况下为 $O(N) + O(N^2 \log N) + O(N)$, 与求最显著边缘时最大特征向量的计算有关, N 为整幅图像的边缘总数. 而 Mahamud 的马尔可夫模型构造的是 $2N \times 2N$ 的状态转移矩阵, 时间复杂度为 $O(N^2) + O(8N^3) + O(N)$ (最优情况下为 $O(N) + O(4N^2 \log(2N)) + O(N)$). 图分割的方法在不限制每个分块尺寸的条件下, 复杂度为 $O(N^e)$, 是 NP 难问题^[2]. 可见, 本文提出的 GSSD 算法有效地降低了运算量, 尤其对较大尺寸的图像, 这一改善更为显著.

4 实验结果

实验环境为 2.4GHz 奔 IV 处理器, 1GB 内存的计算机, 测试图像来自 <http://marathon.csee.usf.edu/PO/Cuts.html> 和 Berkley 图像库. 我们在

4 类共 132 幅图像上进行实验：48 幅室外背景自然目标(class1),28 幅室内背景自然目标(class2),44 幅室外背景人造目标(class3),12 幅室内背景人造目标(class4),图像尺寸均调整为 320×240 。产生最终编组平均耗时 16s。Neut 和 Mini-cut 对 144×96 的图像(在该实验环境下无法在更大尺寸的图像上运行)平均耗时 15s。Neut 和 Mini-cut 是以像素点为对象的编组,故在图像尺寸方面受很大限制,本文方法以边缘为对象的编组,极大降低了待编组数据的维数。

权值根据每幅图像确定,具体方法是对任意一幅图像,提取出所有的闭合轮廓并对每个闭合轮廓分别计算出 SC 和 SP 后,所有 SC 的最大值与除最大值外其他 SC 的均值之差的平方赋值给 α ,用同样的方法计算 β 。对不同的图像,权值是不同的。统计地看,自然景物图像(如动物、人脸等) $\alpha > \beta$,人造目

标图像(如建筑物、雕塑等) $\alpha < \beta$ 。这可能是因为人造目标多具有平行结构,而自然景物多为光滑的闭合曲线轮廓。

以图分割的方法中编组质量最好的 mini-cut^[2]和适用范围最广的 Normalized-cut (Neut)^[4]为代表,比较图分割的方法与 GSSD 算法的编组质量。衡量编组质量的通用标准^[2]

$$\beta = \sqrt{(\sum_{ij} N_{G_i \cap O_j} / N_{G_i})(\sum_{ij} N_{G_i \cap O_j} / N_{O_j}) / N_{ol}} \tag{18}$$

其中, N_{G_i} 是算法产生的编组 G_i 中的边缘总数; N_{O_j} 是真实编组 O_j 中的边缘总数;算法产生的编组与真实编组相重叠的块数为 N_{ol} ;对于每个重叠块 $N_{G_i \cap O_j}$ 是重叠块内的边缘数目。 β 越大意味着错误编组率和漏编率越低。表 1 列出了 GSSD 算法与图分割方法的 β 指标的对比情况。编组的效果如图 3 所示。

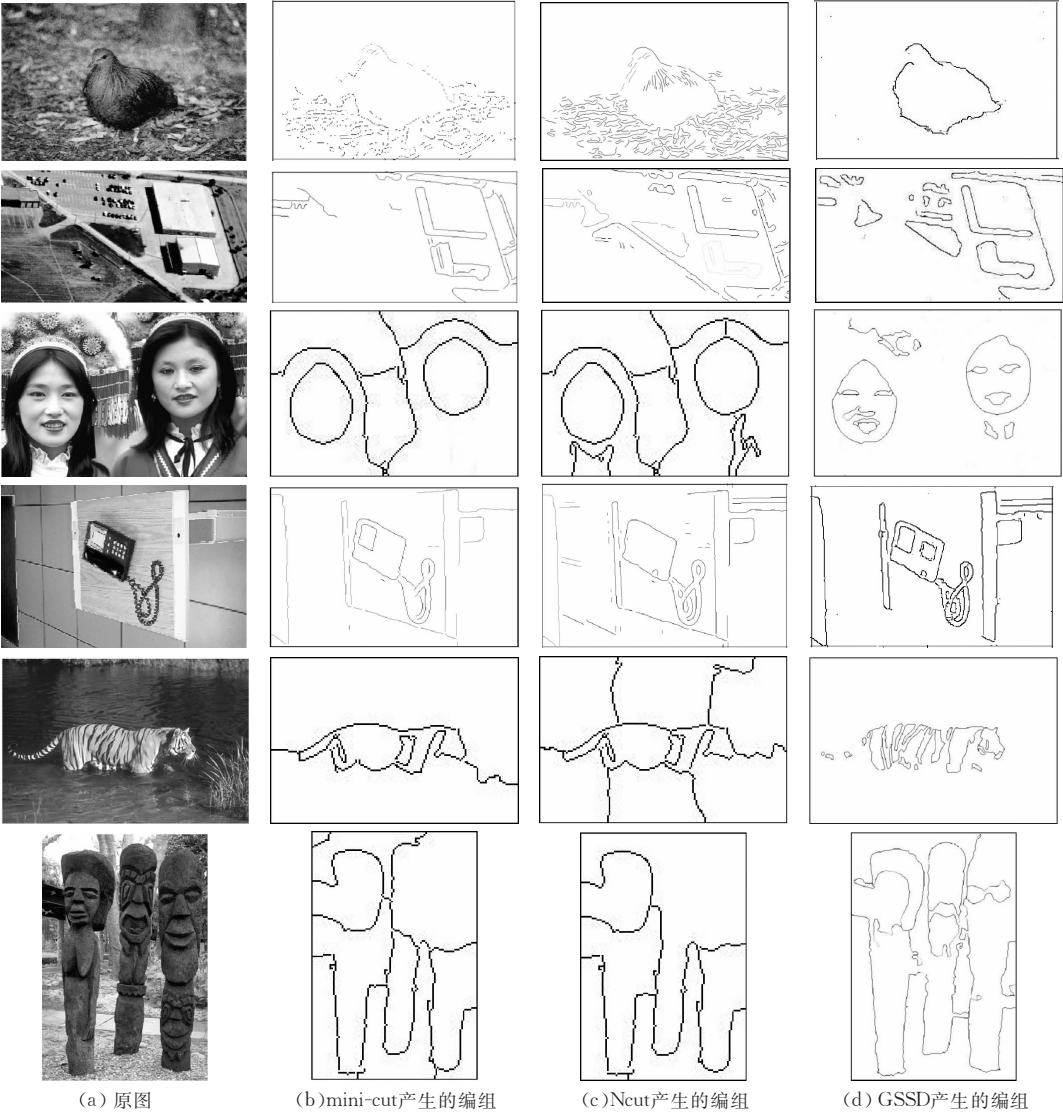


图 3 编组结果

表 1 GSSD 算法与 mini-cut, Ncut 算法的 β 指标的比较

算法	β			
	class1	class2	class3	class4
GSSD	0.61	0.63	0.52	0.43
mini-cut	0.55	0.53	0.48	0.39
Ncut	0.52	0.52	0.49	0.42

可以看出对纹理少的图像编组效果较好,与真实编组基本吻合,而 mini-cut, Ncut 很难将 Canny 提取的背景与目标完全分开. 这是由于本文定义的闭合性这一全局显著结构,将不参与构成目标闭合轮廓的边缘排除在最显著的编组之外,因此足以将目标的完整轮廓从嘈杂背景中提取出来,并作为最显著编组首先被检测到. 对纹理多的目标图像编组效果较差(如图 3 第 5 行),这是仅利用边缘特征的方法所无法克服的问题,mini-cut 和 Ncut 编组时则考虑了灰度、纹理、边缘特征的融合,因此适用范围更广. 根据两个图像库上的测试统计结果,GSSD 算法的编组质量好于图分割的方法,进一步证明我们定义的全局拓扑关系与图分割方法依赖的局部关系相比,提供了更可靠的编组依据.

与按其它方式定义的全局关系相比,本文定义的全局拓扑关系在有遮挡的情况下鲁棒性更好. 如图 4(a),图像背景是有明显木纹的桌面,目标是 3 个水果,有轻微的遮挡. 对这幅图像的编组,Mahamud 的方法(见引言)不能检测到被遮挡的目标对应的编组(图 4(c)),而按本文定义的闭合性能够准确地检测到所有目标(图 4(b)). 因为 Mahamud 的方法要求目标轮廓足够光滑并且组成轮廓的边缘对比度强烈,这幅图像在遮挡处轮廓断裂且对比度较低,不满足 Mahamud 方法对轮廓的约束条件,因此发生漏检. 从统计学的角度来看,以局部关系线性叠加作为全局关系的方法和以二元关系为编组标准的图分割算法本质上都是依据一对边缘间的关系,也就是图像边缘的二阶统计特性实现编组. 本文定义的闭合性不是把一组边缘看成彼此之间独立的边缘对,而是以体现全体边缘间依赖关系的最大特征向量作为全局显著性的衡量指标,反映了边缘分布的整体规律,本质上是依赖高阶统计特性实现编组. 这也是本文定义的全局拓扑关系可靠性更高的内在原因.

需要指出的是,GSSD 算法在构造转移概率矩阵时只考虑了某一条边缘与其最近邻的 6 条边缘之间的转移概率,这对大多数自然图像是适用的,但是当目标被遮挡部分较大时,仅考虑最近邻的 6 条边缘就不能够发现被遮挡部分隔断的、相距较远的边

缘间的关系. 这时需要考虑邻域范围更大边缘间的关系. GSSD 算法失效的另一种情况是背景过于杂乱以至于几乎掩盖了目标轮廓的规律. 如图 4(d)所示,取均匀分布在半径为 16(单位 pixel)的圆周上的小切线段作为待编组的目标,组成目标的切线段条数为 10,将目标置于 64×64 的均匀分布着随机噪声的背景图像的正中,组成背景的切线段条数为 100,每一条背景切线段的位置和朝向均是随机的. 对这幅信噪比为 1 : 10 的图像,GSSD 算法不能产生完整闭合的圆的编组(图 4(e)),而 Mahamud 的方法在这种条件下能产生准确的编组(图 4(f)),这是因为他的方法考虑了边缘的朝向性(每条边缘有相差 180° 的两种朝向,这样形成的闭合轮廓也有顺时针和逆时针两种方向),以两个朝向的交集对应的强连接图确定闭合轮廓,因此在满足一定的约束条件下有更强的抗噪声干扰能力. GSSD 算法如果也考虑边缘的朝向性,其抗噪能力是否能提高,这是我们后继工作将要研究的内容. 但有一点可以确定,考虑两个方向的边缘并利用不对称的转移矩阵,将会增加存储开支和计算复杂度.

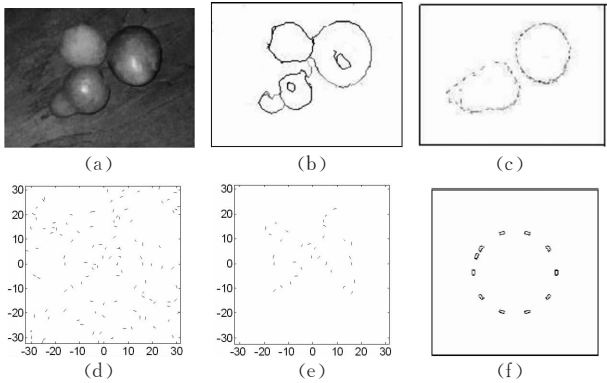


图 4 GSSD 算法(第二列)与 Mahamud 的方法(第三列)编组结果比较((c)和(f)来自文献[11])

5 结束语

提出全局显著结构主导下的知觉编组算法(GSSD 算法),并应用在对不同类自然图像的轮廓提取上. 相对于现有方法的改进主要体现在两个方面:(1)从统计的角度,以转移矩阵的最大特征向量作为全局显著结构的衡量指标,相比于图分割方法中边缘对之间的局部关系和其它方式定义的全局关系(如局部关系之和、马尔可夫过程),该全局显著结构为编组提供了更可靠的依据.(2)模拟注意机制的工作原理,将起点确定在最显著的边缘上并将搜

索空间限制在局部, 这样极大降低了计算复杂度.

将 GSSD 算法应用在公用的编组图像库中, 并用通用的编组质量衡量指标去检验, 结果有了改善. 没有考虑纹理特征影响了该算法对纹理多的图像的编组准确率. 但与融合灰度、边缘、纹理特征的图分割方法的比较数据表明, 该算法仅用到边缘、轮廓的特征, 就产生了与 Ncut, mini-cut 相当的准确率, 进一步证实了全局显著结构在编组中的鲁棒性. 如何定量描述拓扑特征长期以来是研究的难点. 此外, 将该模型应用到更复杂的图像也是我们今后研究的方向.

参 考 文 献

- [1] Palmer S E. Modern Theories of Gestalt Perception. Humphreys G W ed. Understanding Vision. Oxford, U. K.: Blackwell, 1992.
- [2] Soundararajan P, Sarkar S. An in-depth study of graph partition measures for perceptual organization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(6): 642-660
- [3] Elder J H, Zucker S W. Computing contour closure//Proceedings of the 4th European Conference on Computer Vision. Cambridge, UK, 1996: 399-412
- [4] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905
- [5] Herault L, Horaud R. Figure-ground discrimination: A combinatorial optimization approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9): 899-914
- [6] Leordeanu M, Hebert M, Rahul S. Beyond local appearance: Category recognition from pairwise interactions of simple features//Proceedings of the IEEE International Confer-

ence on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 1142-1149

- [7] Liu Jun-Yi, Wang Run-Sheng. Graph representation for objects based on hierarchical decomposition of skeletons. Chinese Journal of Computers, 2001, 24(6): 633-637 (in Chinese)
(刘俊义, 王润生. 基于骨架层次分解的目标的图表示. 计算机学报, 2001, 24(6): 633-637)
- [8] Williams L R, Jacobs D W. Stochastic completion fields: A neural model of illusory contour shape and salience. Neural Computation, 1997, 9(4): 837-858
- [9] Mordohai P. Tensor voting: A perceptual organization approach for computer vision and machine learning//IEEE International Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, Short Course, 2007
- [10] Elder J H, Krupnik A, Johnston L A. Contour grouping with prior models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(6): 661-674
- [11] Mahamud S, Williams L R, Thornber K K, Xu K. Segmentation of multiple salient closed contour from real images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(4): 433-444
- [12] Hoogs A, Collins R, Kaucic R et al. A common set of perceptual observables for grouping, figure-ground discrimination, and texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(4): 458-474
- [13] Wang Zhao-Rui, Shi Rong-Chang. Matrix Analysis. Beijing Institute of Technology Press, 1989 (in Chinese)
(王朝瑞, 史荣昌. 矩阵分析. 北京: 北京理工大学出版社, 1989)
- [14] Itti L, Koch C. Feature combination strategies for salience-based attention system. Journal of Electronic Imaging, 2001, 10(1): 161-169
- [15] Itti L, Koch C. Computational modeling of visual attention. Nature Reviews Neuroscience, 2001, 2(3): 194-203

附 录

引理. 转移矩阵的 n 次幂 P^n 的第 i 个对角元素 p_{ii}^n 表示经过边缘 e_i 的长度为 n 的闭合轮廓的概率.

证明. 首先明确闭合轮廓的概率定义为闭合轮廓上所有相邻边缘间转移概率的乘积, 闭合轮廓至少由 3 条边缘组成. 假设转移矩阵 P 为 $M \times M$ 维, 用递归法证明上述引理.

当 $n=3$ 时, $p_{ii}^3 = (P^3 P)_{ii}$, 因 $p_{ij}^2 = \sum_k p_{ik} p_{kj}$, 故 $p_{ii}^3 = (P^2 P)_{ii} = \sum_j (\sum_k p_{ik} p_{kj}) p_{ji} = \sum_j \sum_k p_{ik} p_{kj} p_{ji}$, 即 p_{ii}^3 等于行程为 $e_i \rightarrow e_k \rightarrow e_j \rightarrow e_i$ ($k, j \in \{1, 2, \dots, M\}$) 的闭合轮廓的概率之和, 也就是经过 e_i 长度为 3 的闭合轮廓的概率;

假设当 $n=t$ 时, 命题成立, 即 p_{ii}^t 等于行程为 $e_i \rightarrow e_{i_1} \rightarrow e_{i_2} \rightarrow \dots \rightarrow e_{i_{t-1}} \rightarrow e_i$ ($i_1, i_2, \dots, i_{t-1} \in \{1, 2, \dots, M\}$) 的闭合轮廓

的概率之和, 也就是经过边缘 e_i 的长度为 t 的闭合轮廓的概率. 类似的, p_{ij}^t 等于行程为 $e_i \rightarrow e_{i_1} \rightarrow e_{i_2} \rightarrow \dots \rightarrow e_{i_{t-2}} \rightarrow e_j$ ($i_1, i_2, \dots, i_{t-2} \in \{1, 2, \dots, M\}$) 的轮廓的概率之和, 也就是起点为 e_i , 终点为 e_j 且长度为 t 的轮廓的概率;

那么当 $n=t+1$ 时, $p_{ii}^{t+1} = \sum_s p_{is}^t p_{si}$, 根据假设 p_{is}^t 表示起点为 e_i , 终点为 e_s 且长度为 t 的轮廓的概率, 故 $\sum_s p_{is}^t p_{si}$ 等于行程为 $e_i \rightarrow e_{i_1} \rightarrow e_{i_2} \rightarrow \dots \rightarrow e_{i_{t-2}} \rightarrow e_s \rightarrow e_i$ ($i_1, i_2, \dots, i_{t-2}, s \in \{1, 2, \dots, M\}$) 的闭合轮廓的概率之和, 也就是 p_{ii}^{t+1} 表示起点为 e_i , 终点为 e_i 且长度为 $t+1$ 的闭合轮廓的概率. 于是, 在 $n=t+1$ 时命题也成立.

证毕.



ZOU Qi, born in 1980, Ph. D. . Her research interests include perceptual grouping, neural computing and image processing.

LUO Si-Wei, born in 1944, Ph. D. , professor. His current research interests include artificial neural network, machine learning, parallel processing, multimedia and Grid computing.

ZHONG Jing-Jing, born in 1981, Ph. D. candidate. Her research interests is computer vision.

Background

The subject is supported by two projects. One is project of National Natural Funds (grant No. 60373029), Effective Coding Model based on Human Vision Perception. The other is project of Department of Education Funds for Doctor (grant No. 20050004001), Research on Mechanism and Algorithms of Perceptual Organization.

The former project adopts the method combining cognitive science and neural science. The target is to advance new cognitive model fitting theoretical framework of effective coding in human vision cognitive system better. Through analyzing cooperation mechanism among multi-vision modules including attention and memory, the authors construct a hierarchical and parallel model, which can approach information processing like human brain to some extent. The results are to be validated by being applied into practical system such as segmentation, grouping and recognition. The latter project researches cognitive mechanism and corresponding computational models of perceptual organization and further applies

the models into intelligent information processing system such as automatic object detection and recognition.

The approach presented in this paper focuses on middle stage of visual information processing-perceptual grouping. Statistical method is employed to organize primitive features from low level into objects, which reflect global structure with scenery meaning. Thus it provides objects for advanced processing in high level. Computation and representation of topological relations are key to grouping algorithms. Existing methods express topological relations through local features or linear combinations of them, which result in bad effect. This paper defines a global topological relation-closure by rigorous probability reasoning and parallelism. So it provides more reliable cues for grouping. Besides, attention is employed to both enhance the model's plausibility in biology way-decided pop-out sequence of groups according to their salience, and improve computational feasibility-reduce complexity.