融合局部与全局信息的头发形状模型

王 楠 [1] 艾海舟 [1]

摘 要 头发在人体表观中具有重要作用, 然而, 因为缺少有效的形状模型, 头发分割仍然是一个非常具有挑战性的问题. 本文提出了一种基于部件的模型, 它对头发形状以及环境变化更加鲁棒. 该模型将局部与全局信息相结合以描述头发的形状. 局部模型通过一系列算法构建, 包括全局形状词表生成, 词表分类器学习以及参数优化; 而全局模型刻画不同的发型, 采用支持向量机 (Support vector machine, SVM) 来学习, 它为所有潜在的发型配置部件并确定势函数. 在消费者图片上的实验证明了本文算法在头发形状多变和复杂环境等条件下的准确性与有效性.

关键词 头发形状建模, 部件模型, 部件配置算法, 支持向量机

引用格式 王楠, 艾海舟. 融合局部与全局信息的头发形状模型. 自动化学报, 2014, **40**(4): 615−623

DOI 10.3724/SP.J.1004.2014.00615

Combining Local and Global Information for Hair Shape Modeling

WANG Nan[1] AI Hai-Zhou[1]

Abstract Hair plays an important role in human appearance. However, hair segmentation is still a challenging problem partially due to the lack of an effective model to handle its arbitrary shape variations. In this paper, we present a part-based model, which is robust to hair shape and environment variations. The model combines local and global information to describe the hair shape. The local model is learned by a series of algorithms, including global shape word vocabulary construction, shape word classifier learning and parameter optimization, while the global model which depicts different hair styles is learned using support vector machine (SVM) to configure parts and define potentials for all underlying hair shapes. Experiments performed on a set of consumer images show our algorithm′s capability and robustness to handle hair shape variations and complex environments.

Key words Hair shape modeling, part-based model, part configuration algorithm, support vector machine (SVM)

Citation Wang Nan, Ai Hai-Zhou. Combining local and global information for hair shape modeling. *Acta Automatica Sinica*, 2014, **40**(4): 615−623

头发在人体表观中发挥着重要的作用, 其分割结果对于人身份识别和图像检索都有巨大的帮助. 认知学的研究[1] 表明头发是人脸识别的重要线索, 其在发型以及表观颜色上的变化会对人脸识别带来很大的误导. 然而虽然头发外观 (颜色纹理等) 已经被用于辅助人脸检索[2], 但发式 (或头发形状) 仍未有广泛使用, 而制约这一点的一个重要因素就是头发在分割上的困难. 同时, 鲁棒的头发分割算法对其他视觉类问题都会有所帮助. 本文算法的 3 种应用: 1) 发型检索, 一方面可以用来搜索与输入图像中人物具有类似发型的图片, 另一方面也可以在限定环境下 (例如证件照中) 搜索具有特定发型特征的人, 可以用于疑犯缉拿中; 2) 虚拟化妆, 可自动将模特发型应用于本人输入图片进行体验; 3) 头发速写, 利用头发分割结果, 辅助人像素描化、动画化.

本文的目的在于建立有效的头发形状模型从而指导头发分割. 需要注意的一点是, 本文中所研究的头发形状指的头发在二维图像所呈现的形状. 同时在本文中, "发型" (Hair style) 指的是典型的头发形状 (Hair shape), 通过对头发形状聚类得到 (几种典型的发型见图 1). 本文采用基于部件的模型, 因为该类模型有能力描述复杂的物体. 为避免歧义, 本文中的部件指的是覆盖物体区域的局部块, 部件的组合构成了对头发形状的粗略估计, 这样设计的原因是很难对头发进行有效的划分, 而类似场景分析类的划分方法在这种情况下更加适用. 对比被广泛研究的刚性物体以及关节型物体, 头发变化更多也更加复杂. 事实上, 头发可以看做是任意非刚性物体, 即其非刚性特征不仅仅在全局上, 而且在所有可能的局部都有. 具体说来, 头发这样的特性给基于部件的模型化方法带来了三个难点. 1) 难以直观地定义 "部件". 对于刚性物体 (例如, 汽车[3] 或者人脸) 或者局部刚性物体 (也被称为关节型物体, 例如牛、

收稿日期 2012-11-02 录用日期 2013-04-19
Manuscript received November 2, 2012; accepted April 19, 2013
国家重点基础研究发展计划 (973 计划) (2011CB302203), 国家自然科学基金 (61075026) 资助
Supported by National Basic Research Program of China (973 Pragram) (2011CB302203) and National Natural Science Foundation of China (61075026)
本文责任编委 章毓晋
Recommended by Associate Editor ZHANG Yu-Jin
1. 清华大学计算机系 北京 100084
1. Department of Computer Science and Technology, Tsinghua University, Beijing 100084

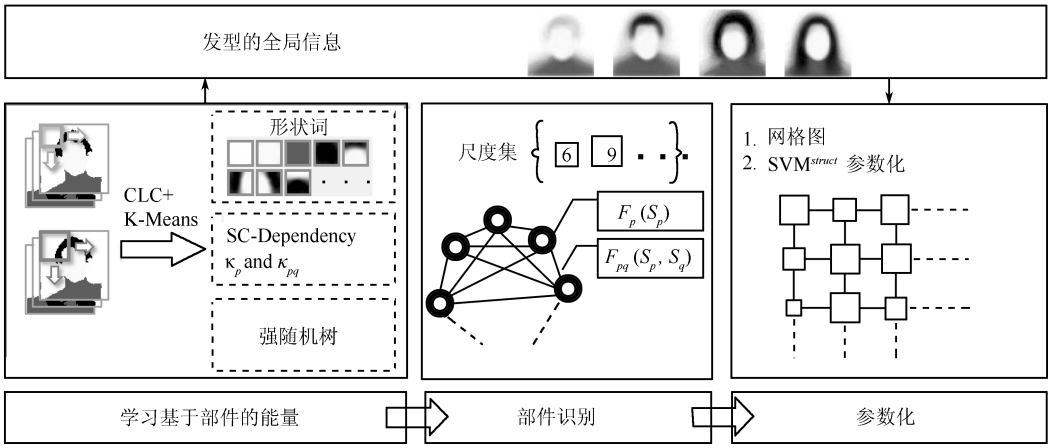


图 1 算法框架流程图

Fig. 1 Illustration of our framework

马[4] 或者人体[5], 部件通常较容易确定 (采取手动或者启发式方法), 也较容易与背景以及其他部件区分开. 然而, 头发在各处都是非刚性的, 很难找到类似的具有分辨性的部件. 即使假设部件可以覆盖头发整体区域将前景背景同时考虑在内, 仍需要合适的部件配置算法, 确保更可靠的部件形状估计. 2) 缺乏清晰有效的部件间约束. 一种可行的方法是根据重叠区域的一致程度[6−7] 来估计. 然而, 这种一致性准则不能刻画长距离约束, 更重要的是它不能保证输出的合理性. 输出合理性在较差光照与较大头发形状变化中具有重要作用, 见图 2. 3) 不同发型形状差别较大, 例如短发与长发, 其部件配置与部件间约束都会非常不同. 这就要求算法自动发现不同的头发形状, 并设定不同的配置与约束.

图 2 本文算法在较大发型变化以及较差光照下的效果

Fig. 2 Results of our algorithm on largely variable hair shapes under bad illuminations

本文的基本思路是将头发形状聚类成典型的发型, 针对不同发型训练更精细的部件模型. 对于不同图片首先判断其所属发型, 继而根据该发型的部件模型, 确定更精细的形状估计. 而算法层面的核心想法是既然头发形状是通过优化部件模型势函数而推理得出, 那么部件的配置可以通过优化对应势函数的有效性来获得. 具体说来, 本文提出了一种新的能量函数, 其能量值对应于部件模型的有效性, 通过优化该能量函数, 便可得到优化的部件配置. 为了达到这一目的, 本文提出了一种可以度量的部件间约束, 称之为子空间聚类约束 (Subspace clustering dependency, SC-Dependency), 并证明 SC-Dependency 保证了模型输出的合理性; 同时, 针对不同发型可能存在不同部件配置的问题, 以部件形状分类器在不同部件配置下的响应作为特征, 设计支持向量机 (Support vector machine, SVM) 分类器, 区分不同发型, 从而得到各自的部件配置与势函数.

算法流程如图 1 所示. 每个部件可能取得的局部头发形状, 称为形状词 (Shape word). 首先, 算法学习一个全局形状词表, 该形状词表是对训练集合中所有可能部件在所有可能尺度上的形状聚类得到. 在这里, 采用正方形的形状词. 相对于为每个部件分别建立一个局部形状词表, 采用全局形状词表不仅可以大大减少重复冗余的局部形状, 还可以增加形状词分类的训练样本, 从而得到更加稳定的、扩展性更好的分类器.

与此同时, 算法对头发整体形状进行聚类得到若干种典型发型 (见图 1). 这些发型被用作全局信息. 在训练阶段, 本文对不同发型生成不同的部件模型. 而为在运行时推导出全局信息, 算法以形状词分类器的响应结果为特征, 训练得到发型分类器. 整体信息提供了对头发形状的粗略估计, 而部件模型则对该形状做更加细致的刻画, 从而使整个模型鲁棒而又能涵盖更多的变化.

之后, 利用已得到的全局与局部信息, 算法通过优化部件模型势函数有效性的方法, 得到每种发型下特定的部件配置结果. 该算法充分考虑了头发的多变性, 在最终的模型中, 同一发型下的不同部件可以具有不同大小, 同时, 同一部件在不同的发型下也

可以具有不同大小.

最终模型采用格点状的马尔科夫随机场描述, 即部件位于覆盖整个头发区域的格点图上, 不同部件具有不同的大小, 而相邻的部件之间 (4 邻域) 会彼此依赖约束, 使得该随机场得到能量最小解的形状词组合构成了最后的头发形状估计 (以概率图形式给出). 利用该形状估计, 结合分割算法, 可以进一步得到像素级头发分割结果.

本文算法在文献 [8] 的基础上进行了改进, 首先, 将全局信息整合到原始模型中, 可以更好地刻画差别较大的发型; 另外, 对于新引入的全局信息, 设计了新的学习算法, 该算法为全局信息与布局信息模型分别学习不同的区分性模型; 最后, 为简化模型, 减少内存与计算消耗, 整体流程尽可能相互依赖, 利用已有结果, 例如使用全局词表, 将词表分类器结果用于发型分类等.

为简明起见, 在接下来的章节中, 我们会先介绍不包含全局发型信息的部件模型及其训练算法. 最后, 会阐述如何对此模型进行扩展, 融入全局信息.

1 相关工作

头发分割是近些年来逐渐受到重视的一个问题. Yacoob 等[9] 根据预先定义的头发位置建立头发颜色模型, 而后迭代更新颜色模型与头发形状并得到最终头发检测结果. 进一步, 针对他们的区域增长算法, Rousset 等[10] 引入了频域信息来增强该模型. 由于头发颜色变化很大, 同时容易与周围环境光照融合而模糊边界, 所以进一步的工作[11−12] 又引入了形状信息来约束头发形状. Scheffler 等[13] 又引入了更加精巧的颜色模型, 分别模型化头发、皮肤、衣服以及背景的颜色分布. 这几种方法都采用了相对简单的形状模型, 使得他们对于较差光照下复杂的头发形状描述能力不强. Wang 等[7] 通过采用基于部件的模型来解决该问题, 他们的方法可以针对不同图像生成自适应的不同形状先验来指导分割. 然而由于采用了搜索的方法来为不同部件寻找模板, 他们的算法计算负担较重.

基于部件的模型被广泛地应用于物体分割中. 其中一些工作[6, 14] 学习具有区分性的片段来估计物体形状而另外的工作[4−5] 采用图案结构模型 (Pictorial structure) 提供高层物体形状信息. 这些算法都是对关节型物体 (Articulated object) 进行建模, 对这类物体相对容易找到 (通过搜索或学习) 具有区分性的部件, 但头发是完全非刚性物体, 无法满足关节型物体的假设.

在分割之外, 基于部件的模型也被用于检测[15], 其中, 不同点之一是在我们的问题中, 部件需要能覆盖整个物体, 即不仅需要模型化具有区分性的部件, 也要模型化不具有区分性的部件. 虽然 Levin 等[16] 探讨了采用稀疏片段对关节型物体进行分割的方法, 但对于完全非刚性物体, 其性能并不能确定.

另外一个相关的工作[17] 中, 物体被递归的分解为局部部件, 并且所有部件共享同一个全局形状词表. 本文工作与该工作有两个主要不同点: 1) 在该论文的工作中, 所有部件在所有层次都采用固定的大小而本文中部件可以进行配置具有不同的大小; 2) 该论文中形状词表是人工手动生成的, 而本文中形状词表是在训练数据上自动学习出来, 不需要昂贵繁重的人工干预.

2 基于部件的头发形状模型

本文的目的是给定一张图片 I, 推理出每个部件的局部形状, 从而可以为头发形状分割提供先验信息. 令 P 表示所有部件的集合, 对于每一个部件 $p \in P$ 定义一个形状词标签 $v_p \in \{1, \cdots, |V|\}$, 其中 V 是所有部件共享的形状词表.

本文的头发形状模型采用马尔科夫随机场进行形式化, 其中部件对应于节点, 而相关的部件在图中有边相连. 给定输入图片以及部件配置, 其部件标注的能量 (损失) 函数可以形式化为如下式:

$$E_v(\boldsymbol{v}|\boldsymbol{x}, \boldsymbol{s}) = \sum_p \theta_p f_p\left(v_p|x_p, s_p\right) + \sum_{pq\in\mathcal{E}_v} \theta_{pq} f_{pq}\left(v_p, v_q|s_p, s_q\right) \tag{1}$$

其中, $\boldsymbol{x}$ 是输入图像特征; $\boldsymbol{s}$ 是部件配置, 即每个部件的大小; 每个 $v_p \in \{1, \cdots, |V|\}$ 对应于赋予部件 p 的形状词; $\theta = \{\{\theta_p\}, \{\theta_{pq}\}\}$ 代表对应势函数项的权重, 衡量每个势函数项对最终结果的影响程度, 该参数采用结构 SVM[18] 学习得到; $\mathcal{E}_v$ 代表所有相关联的部件.

一阶势函数 $\{f_p\}$ 衡量将每个部件的图像特征映射为不同形状词的损失, 在本文中我们采用随机森林[19] 的形式表达. 二阶势函数 $\{f_{pq}\}$ 对相关部件上两个形状词同时出现的概率进行建模.

3 形状词表与子空间聚类约束

形状词表是本文模型的核心部分, 并且可以自然地引出一种可度量的约束, 即子空间聚类约束. 本节示意性说明子空间聚类约束的思想, 并说明即使采用全局形状词表, 子空间聚类约束仍然可以保证输出合理性.

3.1 子空间聚类

假设需要为一组几何形状建立部件模型, 见图 3. 训练数据在不同子空间上 (左侧形状和右侧形状)

分别被聚成不同的类别, 聚类结果分别用虚线与实线标记出来, 每个部件分别建立词表. 根据子空间聚类的结果, 两个形状词同时发生的概率 κ 可以通过统计对应聚类结果重叠度来得到, 而这也为算法从局部恢复整体提供了统计约束.

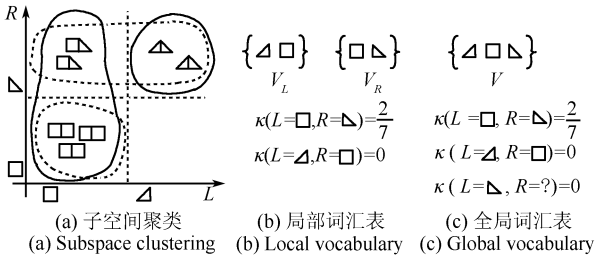


图 3 说明子空间聚类约束的例子

Fig. 3 Toy example to explain SC-Dependency

子空间聚类约束的一个重要特性就是任何两个形状词 $v_l \in \{1, \cdots, |V_L|\}$ 与 $v_r \in \{1, \cdots, |V_R|\}$, 它们同时出现当且仅当 $\kappa(v_l, v_r) > 0$. 这样当把子空间聚类约束引入到能量函数 (1) 中时, 就可以保证在最优解中出现的形状词的组合一定是合理的, 即任何两个在最优解中同时出现的形状词一定在训练集中至少共享一个训练样本. 合理性对部件模型具有很重要的意义, 它保证了算法在较差光照条件下的稳定输出, 如图 4 所示. 从图 4 中可以看出, 文献 [7] 中算法的输出具有明显的不合理性, 而本文算法保证即使在边界相对模糊的区域仍能输出合理的头发形状, 即根据邻域信息预测边界的位置, 更多复杂环境下的形状估计结果见第 6.4 节.

进一步观察可以发现, 不同部件的词表可能共享某些形状词 (例如图 3 (b) 中的正方形), 在实际问题中, 这种现象非常普遍. 所以相对于为每个部件分别建立局部形状词表, 本文为所有部件建立共享的全局形状词表. 而在全局形状词表中, 之前提到的输出合理性仍然可以保持, 如图 3 所示. 另外, 采用全局形状词表不仅会减少冗余的形状词, 而且可以为形状词分类的训练提供更多的数据, 从而得到更稳定可靠的分类器. 这一点对于训练数据相对较少的分割类问题尤为重要.

给定 N 个形状, 通过对所有部件的局部形状同时聚类得到全局形状词表 V. 令 $\pi_{p,n}$ 为第 n 张图片上的第 p 个部件形状的聚类标签 (即形状词索引), 则一阶与二阶子空间聚类约束可以通过如下式计算:

$$\kappa_p(v_p) = \frac{\sum_n \delta(\pi_{n,p} = v_p)}{N} \tag{2}$$

$$\kappa_{pq}(v_p, v_q) = \frac{\sum_n \delta(\pi_{n,p} = v_p)\,\delta(\pi_{n,q} = v_q)}{N} \tag{3}$$

其中, $v_p \in \{1, \cdots, N\}$ (或 v_q) 为部件 p (或 q) 的词索引, 而 δ 为指示函数.

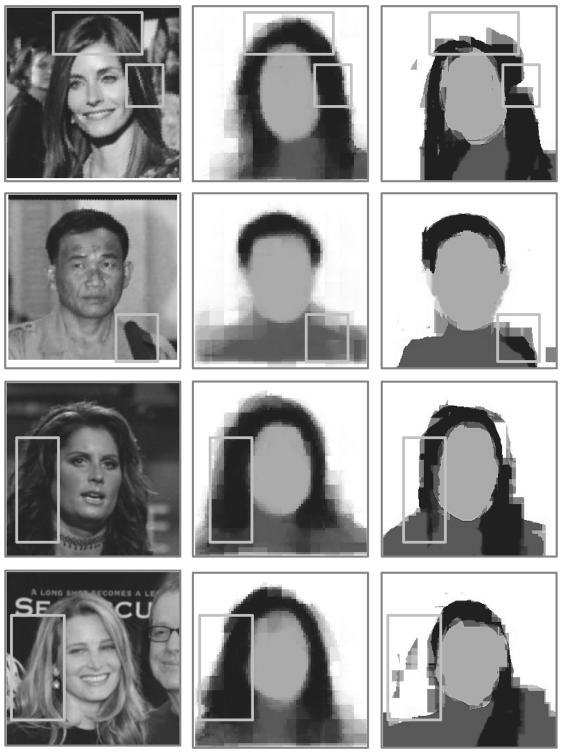


图 4 与文献 [7] 中算法对比, 说明输出合理性的重要性

Fig. 4 Importance of output reasonability compared with the method in [7]

式 (1) 中的二阶势函数则定义为相关部件的子空间聚类约束的负对数:

$$f_{pq}(v_p, v_q) = -\log \kappa_{pq}(v_p, v_q) \tag{4}$$

对于非法的词组合, 子空间聚类约束为零. 此时, 其二阶势函数定义为一个足够大的数字 (本文的实验中采用 10E6), 从而保证他们不会同时出现在最优解中.

3.2 建立形状词表

本文中, 形状词表通过对所有部件在所有尺度上的局部形状进行聚类得到. 这里所有尺度取自一个离散的尺度集合 S, 详细参数请见实验.

聚类时, 所有局部形状下采样到相同尺度, 并采用汉明距离计算形状之间的不相似程度. 算法的目标是希望所得到的形状词表能够以最小误差近似原始数据, 所以比较合理的选择是采用 K-Means 算法进行聚类. 但是, 在本文的问题中, 因为形状词的分布差异很大 (图 5), K-Means 算法常常因为其随机初始化而失效. 为了解决这一问题, 我们首先采用凝聚型聚类算法进行初步聚类, 而后用此聚类结果

初始化 K-Means 算法, 从而得到最终的聚类结果. 实验证明, 采用该初始化方法的 K-Means 算法, 其性能要远远好于随机初始化的 K-Means 算法, 也同样好于只采用凝聚型聚类算法 (约有 3.8 % 的提升), 这里形状词在约 2×10^5 个局部形状上聚类得到, 性能采用类内方差和来衡量.

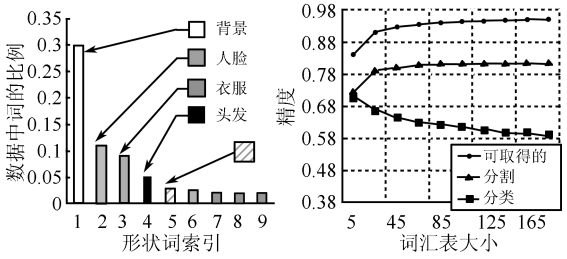


图 5　形状词表分析

Fig. 5　Analysis of shape word vocabulary

另一个问题是词表的大小. 通过实验 (见图 5 准确率曲线[1]) 可以观察到, 虽然随着词表规模增大, 其分类正确率在逐步降低, 但是其在分割上的准确率仍然稳步提高. 本文的实验采用 $|V| = 120$ 的词表.

4　部件配置

部件配置对于头发形状模型具有重要意义, 见图 6. 头发的局部信息有时具有歧义, 不能对形状分类提供足够的信息. 更大的头发部件可以减少歧义, 但这并不意味着头发部件越大越好, 一个直观的解释是小的部件更容易产生紧致的局部形状分布, 从而更容易进行形状词分类. 另外, 对有些部件, 局部信息并不足以确定局部头发形状, 这时候往往需要其他部件的辅助来完成分类. 本文的一个重要贡献就是提出了一种新的部件配置算法, 该算法通过直接提升头发部件模型的有效性来确定最优的头发部件配置.

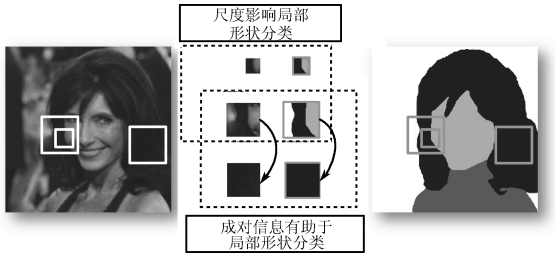


图 6　部件配置的重要性

Fig. 6　Importance of part configuration

[1]其中, 可取得的准确率根据聚类带来的误差计算, 是算法可能取得的最高准确率; 分类准确率指将不同形状词当成完全不同类别得到的准确率; 分割准确率指将不同形状词根据其相似程度计算误差而得到的准确率, 可以更准确地体现本文的需求.

假设每个部件的一阶势函数 f_p 在尺度 s_p 上的有效性记为 $F_p(s_p)$; 每一对相关部件之间的二阶势函数 f_{pq} 在尺度 (s_p, s_q) 上的有效性记为 $F_{pq}(s_p, s_q)$, 其中, F_p 和 F_{pq} 都满足值越小, 对应的解越优. 那么, 部件配置的最优解可以通过最小化如下能量函数来得到:

$$E_s(\boldsymbol{s}) = \sum_p F_p(s_p) + \lambda \sum_{pq \in \mathcal{E}_s} F_{pq}(s_p, s_q) \qquad (5)$$

其中, $\mathcal{E}_s$ 与式 (1) 中的 $\mathcal{E}_v$ 不同, 在本文实验中, $\mathcal{E}_s$ 构成完全图, λ 是二阶权重.

通常一阶势函数 f_p 都是通过机器学习的算法来确定 (例如 Boosting 算法[20] 或随机森林算法[21], 则其有效性可以通过其测试准确率 (交叉验证准确率或 Out-of-bag 准确率) 来得到, 即 $F_p(s_p) = -\log(\mathrm{Ac}_{\mathrm{oob}}(p, s_p))$, 其中, $\mathrm{Ac}_{\mathrm{oob}}$ 是随机森林的 Out-of-bag 准确率.

二阶势函数, 比如重叠区域的一致性, 通常采用启发式方法来定义, 而它们的有效性比较难以直观衡量. 本文采用子空间聚类约束来定义部件间约束, 而该约束可以在信息论框架下衡量其有效性:

$$F_{pq}(s_p, s_q) = \exp \frac{-I_{pq}(s_p, s_q)}{\beta} \qquad (6)$$

其中, $I_{pq}(s_p, s_q)$ 是部件 p 与 q 在 s_p 与 s_q 的部件配置下的互信息. 它的值具体是通过约束在相关部件尺度上的子空间聚类约束 $\kappa(v_p, v_q | s_p, s_q)$ 来估计. β 是归一化因子, 定义为所有互信息的期望.

优化如式 (1) 的全连接概率图是 NP-hard 问题. 本文采用 TRW-S 算法[22] 来近似求解.

5　发型分类

本文之前章节的算法已经建立了一个一般性的头发形状模型. 该模型的不足之处在于缺少对于全局信息的衡量. 虽然马尔科夫随机场被用来关联不同的部件, 但对于发型整体的刻画并不包含在其中, 这也是马尔科夫随机场的缺陷之一. 为了解决这一问题, 本节将对头发形状的全局信息进行建模, 其基本想法是通过对头发整体形状进行聚类, 得到典型的发型, 再将训练数据划归到不同的发型中, 继而为每一类发型分别统计子空间聚类约束. 通过约束训练数据, 每一类发型部件模型的二阶势函数的有效性 (见式 6) 都可以得到很大提升, 从而大大提高模型的准确性.

在此情况下, 头发形状模型可改写为

$$E_v(\boldsymbol{v}|h,\boldsymbol{x},\boldsymbol{s}) = \sum_p \theta_p f_p\left(v_p|h,x_p,s_p\right) + \sum_{pq\in\mathcal{E}_v} \theta_{pq} f_{pq}\left(v_p,v_q|h,s_p,s_q\right) \tag{7}$$

其中, 不同之处在于增加的 h 变量, 它被用来表示本图片中的头发整体上应属的发型类别. 在训练时, 所属类别是通过对头发形状的相似程度来确定, 但在测试时, 头发形状为所求. 本文采用了中间语义层的特征来完成发型的分类. 具体说来, 首先根据全局形状词表与发型的聚类结果, 统计每个形状词属于各发型的概率. 对于每一张图片, 通过形状词随机森林分类器, 计算其每个部件, 在每个尺度上对每个形状词的响应, 并将其对各发型的概率组成特征向量, 继而采用 SVM[23] 训练得到分类器. 这里每一个响应代表了图片每一个局部的特征, 它们的组合将从全局角度刻画头发形状信息, 从而弥补此前模型的欠缺.

6 实验

本文在 Labeled Faces in the Wild 集合[24] 上进行实验. 该数据集中的每张图片包含有至少一张人脸, 并截取了足够大的部分使得头发区域也包含于其中, 在本文的实验中, 只对每张图片中的主要人脸的头发部分进行估计. 这些图片都是从网络收集而来, 所以包含了比较大的头发形状变化, 也包含了复杂环境与较差光照条件. 我们手工标定了 1 021 张图片, 图片中每个像素分别标记成为 4 个标签之一, 分别是脸部、头发、肩部以及背景. 每次实验随机挑选一半数据用以训练, 另一半数据用以测试, 此过程重复 5 次, 最终平均结果作为性能统计结果.

6.1 实现细节

所有图片根据人脸检测的结果, 规范化到 72 像素 × 72 像素大小, 其中人脸大小为 24 像素 × 24 像素大小. 所有的部件都是正方形, 其可能取得尺度为 $\{6,9,12\}$ 像素. 部件模型中边集 $\mathcal{E}_v$ 按照格点图来建立联系, 其中每 6 个像素设置一个节点, 即共有 144 个节点. 全局发型根据汉明距离共聚为 4 类. 学习过程如下:

学习算法流程

1) 对局部形状聚类得到全局形状词表, 并采用随机森林学习形状词分类器;

2) 对全局形状聚类, 得到整体发型, 并统计形状词属于各发型的概率;

3) 对每幅图像, 根据 1), 2) 计算各尺度部件属于各发型类别的概率, 以此为特征使用 SVM 训练发型分类器;

4) 分别对每个发型, 统计 SC-Dependency, 优化式 (5), 配置部件大小, 并采用结构 SVM[18] 优化式 (1) 中的参数.

SVM[23] 中的 C 与 γ 采用交叉验证得出. 对于结构 SVM[18], 其标注结果之间的间隔 δ 根据形状词之间的汉明距离来计算; 规范化参数 C 采用交叉验证得出; 算法结束参数 ϵ 为 0.05; 其他参数采用默认值. 部件配置算法中的权重 λ 通过交叉验证统计得出, 本文中采用 $\lambda = 0.12$. 对于全局信息, 整体发型 (根据整体汉明距离) 聚为 4 类.

实验中, 形状词分类器采用随机森林学习得到, 其训练数据为所有部件在所有尺度上的局部形状. 这些图像块被规范化到相同大小, 并被转化到 CIELab 颜色空间, 而后与 Semantic Texton Forest[21] 类似, 抽取如下特征: 1) 位于 (x,y) 位置的单个像素在通道 b 上的值 $I_{x,y,b}$; 2) 一对像素值之和 $I_{x_1,y_1,b_1} + I_{x_2,y_2,b_2}$; 3) 一对像素差值 $I_{x_1,y_1,b_1} - I_{x_2,y_2,b_2}$; 4) 一对像素差值的绝对值 $|I_{x_1,y_1,b_1} - I_{x_2,y_2,b_2}|$. 同时, 因为不同部件所处位置不同会大大影响其分类结果, 我们又提出了特征 (5), 即该像素离图像中心的规范化距离 $\sqrt{\frac{(x-w/2)^2+(y-h/2)^2}{(w/2)^2+(h/2)^2}}$. 另外, 鉴于形状词间相似程度并不相同, 在计算决策树的节点分裂策略时, 本文算法选取能产生最大一致度的分裂决策. 其中对每一组数据 D, 其形状一致度 $C(D)$ 可以通过下式算得:

$$C(D) = \frac{\sum\limits_{i,j\in D} w_i w_j c\left(v_i,v_j\right)}{\sum\limits_{i,j\in D} w_i w_j} \tag{8}$$

其中, $c(v_i,v_j)$ 是形状词 v_i, v_j 之间的相似程度, 而 w_i, w_j 为数据点的权重. 注意到我们只需要统计形状词之间的距离, 而不需要真的计算每两个局部形状的距离, 而形状词相对较少 (相较于约 30 000 的局部形状), 所以通过精巧的调整计算顺序, 上式可以在 $\mathrm{O}(N)$ 时间内计算出, 从而保证了本文算法的高效性.

具体实现时, 随机森林由 20 棵决策树组成, 每个节点进行 5 000 次分裂检测, 从中选取最好的分裂方式, 每棵决策树使用 1/4 的数据进行训练, 并且决策树不进行任何剪枝操作.

6.2 形状词分类

随机森林在本文算法中发挥重要作用, 本节实验将比较本文中随机森林方法与传统 Semantic Texton Forest[21] 的性能. 本文模型将用于头发分割, 所以我们更关心模型在分割上的正确率. 在图 7 中, 准确率是通过形状词间的相似程度来计算. 可以看出, 特征 5 (即相对位置特征) 可以带来明显的性能提升 (约 17 %). 另外我们统计了不同特征在决策

树中出现的平均深度, 发现特征 5 出现的平均深度约为 2, 而其他特征都在 3 以上, 即特征 5 出现在决策树中更靠近根节点的层次, 这就说明位置特征具有更强的区分能力; 另外, 我们还统计了不加入与加入特征 5, 决策树规模 (节点个数) 的差异, 统计表明, 加入特征 5 可以将决策树规模减少约 40 %, 从而减少内存与计算消耗.

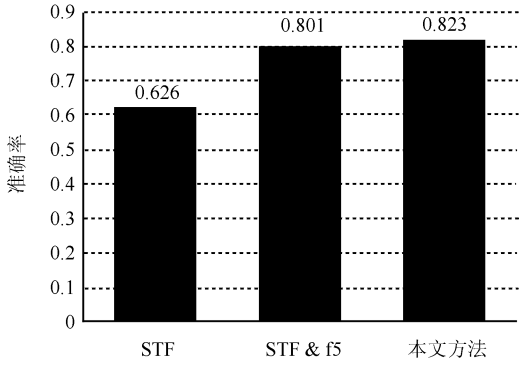


图 7　形状词分类器性能比较

Fig. 7　Accuracy of shape word classification

同样, 实验也表明修改后的节点分裂准则也可以在一定程度上提升随机森林的性能 (约 2 %). 这是因为新的节点分裂准则并不是简单地希望分裂后的两组数据各自包含尽可能紧致的形状词分布, 而是更倾向于将相似的形状词放在同一组数据中. 值得注意的是, 由于涉及到的像素数目极多 (约 10E6), 因此, 新的节点分裂准则也带来显著的分割结果上的改善.

6.3　部件配置优化

本节通过与使用不同部件配置的模型比较来研究本文提出的部件配置算法的性能. 实验结果见表 1, 其中准确率通过像素级分割准确率来计算. "单一尺度" 的性能将只采用一个尺度的模型性能平均后得到; "多尺度" 的性能是在同一模型中采用所有尺度的部件来确定的性能; "优化尺度" 即本文算法得到的模型所取得的性能. "多尺度" 与 "优化尺度" 都对 "单一尺度" 方法有明显的提升. 然而 "多尺度" 方法包含了更多的歧义部件, 这些部件的错误有时候会扩散到其他正确的部件中, 从而带来性能的降低. 本文的 "优化尺度" 将部件的歧义考虑在内, 可以取得更好的性能.

表 1　不同部件配置的性能比较

Table 1　Comparison of different part configurations

	单一尺度	多尺度	优化尺度
准确率 (%)	84.3	86.2	91.7

6.4　与其他方法的对比

本节首先将本文算法与文献 [7] 中算法进行了对比, 见表 2. 在这里, 表格中所有数据针对头发像素级别的分割结果进行衡量. 数据表明, 本文中的算法比文献 [7] 中的算法在分割性能上有一定提升, 同时取得近 80 倍的速度提升. 这种速度上的提升, 主要来源于本文采用了全局形状词表与形状词分类器的技术, 从而使得文献 [7] 中 O(N) 的搜索时间 (其中 N 为标注好的图片数量) 缩减为 O(1) 的分类时间. 需要注意的是, 全局形状词表的大小会影响本文算法的具体运行时间, 但标注库大小的变化不会直接影响运行时间. 而在具体的定性实验中, 可以发现, 本文的算法对于复杂环境与较差光照下的头发分割具有更加鲁棒的效果, 见图 4.

表 2　与文献 [7] 中方法的对比

Table 2　Comparison with the method in [7]

	FAR (%)	FRR (%)	时间 (s)
算法 [7]	6.2	24.7	131.2
本文算法	3.0	20.1	1.7

其次, 我们还比较了本文算法与 Texton boosting (TB)[20]、Adapted color model (AC)[13] 和 MRF-based model (MRF)[11]. 其中前二者代码由作者主页获得, 第三种算法代码为自行实现, 三者使用模型均由本文标注数据重新训练得到. 在这里, 为适应不同算法要求, 性能只通过头发的像素级分割结果进行评测. 其结果如下表 3 所示.

表 3　与其他算法方法的对比

Table 4　Comparison with other methods

	本文算法	TB	AC	MRF
准确率 (%)	93.1	56.2	88.4	90.7

从表 3 中数据可见, Texton boosting 算法 (使用了约 700 轮 Boosting 迭代) 虽然在纹理分类上效果很好, 但在头发的分割方面表现不佳. 一个可能的原因是在头发分割中, 背景环境过于混乱, 此时纹理信息不如位置信息有分辨性. Adapted color model 与 MRF-based model 也都给出了相对接近的结果, 但 MRF-based model 性能更好一些, 其原因在于后者使用了多个形状模板, 对头发形状信息有更好的刻画. 本文算法一方面在发型的分类上采用了机器学习的方法, 比 MRF-based model[11] 更加准确; 另一方面在每种发型内部, 仍能进行微调, 使得生成的模板更加贴合特定输入图像, 从而性能更优.

我们在图 8 中测试了本文算法在较为复杂环境下的头发形状估计结果. 这些图片中, 环境中存在与

头发颜色接近的物体, 头发与环境的部分边界并不清晰. 本文算法结合了全局与局部的线索, 利用具有较强分辨性的块通过子空间聚类约束的输出合理性约束, 保证即使存在模糊边界, 模型输出仍然稳定可靠.

6.5 定性分割

本节实验展示了将本文算法应用于头发分割框架所得到的最终结果. 如图 9 所示, 本文算法对头发形状变化和环境光照变化都有较鲁棒的输出结果. 在具体分割时, 本文算法输出的头发概率模板作为一阶势函数之一加入到分割能量函数中, 同时在分割框架中还加入了表观模型 (RGB 直方图) 以及超像素约束. 可以看出由于本文算法提供的鲁棒的形状模型的帮助, 对于即使光照较差条件下的图片, 最终的分割算法仍能准确找到物体边界, 例如与黑色环境相互融合的头发边界区域.

图 8 复杂环境下的头发形状估计

Fig. 8 Hair shape estimation in complex environments

图 9 定性的头发分割结果

Fig. 9 Qualitative hair segmentation results

7 结论

本文提出了一种用于头发形状建模的部件模型. 该模型由可配置的部件组成, 可以为每一个部件取得最适应的尺度; 同时, 也融合了全局发型的信息, 使得模型对不同发型具有更好的针对性. 在模型建立的过程中, 本文提出了子空间聚类约束, 并将其融合到模型框架中. 子空间聚类约束一方面可以保证部件模型输出的合理性, 从而使得模型在复杂环境与光照条件下仍能提供鲁棒的形状信息; 另一方面也提供了二阶势函数的信息论有效性度量方法, 从而协助部件配置算法搜索最优部件配置. 最终, 我们在非常具有挑战性的消费者图片上进行了详尽的实验, 证明了本文模型的优良性能.

References

1 Sinha P, Poggio T. United we stand: the role of head structure in face recognition. *Perception*, 2002, **31**(1): 131−133

2 Kumar N, Belhumeur P N, Nayar S K. Facetracer: a search engine for large collections of images with faces. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer-Verlag, 2008. 340−353

3 Winn J, Shotton J. The layout consistent random field for recognizing and segmenting partially occluded objects. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 37−44

4 Kumar M P, Torr P H S, Zisserman A. OBJCUT: efficient segmentation using top-down and bottom-up cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(3): 530−545

5 Andriluka M, Roth S, Schiele B. Pictorial structures revisited: people detection and articulated pose estimation. In: Proceedings of the 2009 Computer Society Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 1014−1021

6 Borenstein E, Ullman S. Combined top-down/bottom-up segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(12): 2019−2125

7 Wang N, Ai H Z, Lao S H. A compositional exemplar-based model for hair segmentation. In: Proceedings of the 10th Asian Conference on Computer Vision. Queenstown, New Zealand: Springer-Verlag, 2010. 171−184

8 Wang N, Ai H Z, Tang F. What are good parts for hair shape modeling? In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 662−669

9 Yacoob Y, Davis L S. Detection and analysis of hair. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(7): 1164−1169

10 Rousset C, Coulon P Y. Frequential and color analysis for hair mask segmentation. In: Proceedings of the 15th IEEE International Conference on Image Processing. San Diego, California, USA: IEEE, 2008. 2276−2279

11 Lee K C, Anguelov D, Sumengen B, Gokturk S B. Markov random field models for hair and face segmentation. In: Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition. Amsterdam, The Netherlands: IEEE, 2008. 1−6

12 Wang D, Shan S G, Zeng W, Zhang H M, Chen X L. A novel two-tier Bayesian based method for hair segmentation. In: Proceedings of the 16th International Conference on Image Processing. Cairo, Egypt: IEEE, 2009. 2401−2404

13 Scheffler C, Odobez J M. Joint adaptive colour modelling and skin, hair and clothes segmentation using coherent probabilistic index maps. In: Proceedings of the 22nd British Machine Vision Conference. University of Dundee: BMVA Press, 2011. 53. 1−53. 11

14 Borenstein E, Malik J. Shape guided object segmentation. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2006. 969−976

15 Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. In: Proceedings of the 2008 Computer Society Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008. 1−8

16 Levin A, Weiss Y. Learning to combine bottom-up and top-down segmentation. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 581−594

17 Zhu L, Chen Y H, Lin Y, Lin C X, Yuille A. Recursive segmentation and recognition templates for 2D parsing. In: Proceedings of the 22nd Annual Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: Curran Associates, Inc., 2008. 1985−1992

18 Tsochantaridis I, Hofmann T, Joachims T, Altun Y. Support vector machine learning for interdependent and structured output spaces. In: Proceedings of the 21st International Conference on Machine Learning. New York, NY, USA: ACM, 2004. 104−111

19 Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Machine Learning*, 2006, **63**(1): 3−42

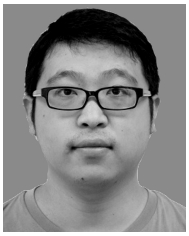
20 Shotton J, Winn J, Rother C, Criminisi A. Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer-Verlag, 2006. 1−15

21 Shotton J, Johnson M, Cipolla R. Semantic texton forests for image categorization and segmentation. In: Proceedings of the 2008 Computer Society Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008. 1−8

22 Kolmogorov V. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(10): 1568−1583

23 Chang C C, Lin C J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, **2**(3): Article No. 27 1−27

24 Huang G B, Ramesh M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, USA, 2007



王 楠 清华大学计算机科学与技术系博士研究生. 2008 年获得清华大学学士学位. 主要研究方向为模式识别和图像语义分割.
E-mail: aaron.nan.wang@gmail.com
(**WANG Nan** Ph. D. candidate in the Department of Computer Science and Technology, Tsinghua University. He received his bachelor degree from Tsinghua University in 2008. His research interest covers pattern recognition and semantic image segmentation.)

艾海舟 清华大学计算机科学与技术系教授. 1985 年、1988 年和 1991 年分别获得清华大学学士、硕士和博士学位. 1994 年 9 月至 1996 年 8 月在比利时布鲁塞尔自由大学做博士后研究. 主要研究方向为计算机视觉. 本文通信作者.
E-mail: ahz@mail.tsinghua.edu.cn
(**AI Hai-Zhou** Professor in the Department of Computer Science and Technology, Tsinghua University. He received his bachelor, master, and Ph. D. degrees from Tsinghua University, in 1985, 1988, and 1991, respectively, as a post doctoral researcher during September 1994 ∼ August 1996 in Brussels University, Belgium. His main research interest is computer vision. Corresponding author of this paper.)