

感知归类在目标识别中的应用研究^{*})

魏丽 吴中福 李云 古毅
(重庆大学计算机学院 重庆 400044)

摘要 感知归类是指人们在不拥有任何图像内容知识的前提下,从底层原始图像特征中提取有效的图像联系的一种视觉能力。本文着重介绍了感知归类方法在目标识别中的应用,通过使用感知归类中的接近、相似、连续等概念,逐步将底层的图像特征组合为高层的图像联系,直至获得有意义的语义表征。

关键词 感知归类,目标识别,特征提取

Application Research on Perceptual Grouping in Object Recognition

WEI Li WU Zhong-fu LI Yun GU Yi

(College of Computer, Chongqing University, Chongqing 400044)

Abstract Perceptual grouping refers to the human visual ability to extract significant image relations from lower-level primitive image features without any knowledge of the image content. In this paper, we introduce the application of perceptual grouping in object recognition. The grouping process hierarchically groups these relations continuously until a meaningful semantic representation is achieved that may be used by a higher-level reasoning process. The grouping principles embodied such concepts as grouping by proximity, similarity, continuation.

Keywords Perceptual grouping, Object recognition, Feature extracting

1 引言

人们对视觉图像的认知,并非是图像中一些各自独立的部分,而是经过感知系统组织后的形态与轮廓的集合。例如人们在欣赏一幅图画或一张摄影作品时,画面里的每一部分形成了各自独立的视觉元素,如果想让观众留下深刻的视觉认知,元素与元素之间必须有某种形式的关连。人类的认知系统对这种关连形式有着极强的敏感性,能够把原本各自独立的局部信息连接整合成一个整体概念^[1]。

2 原理简介

感知归类是指人们在没有关于图像内容任何知识的前提下,从底层的原始图像特征中提取有效的图像关系的视觉能力。感知归类从图像的底层特征开始,逐步提取图像的高层特征,直至获得有意义的语义表示,该语义表示可用于更高层次的推理过程。由 Gestalt 提出的归类原则包含了以下的概念:接近性,相似性,连续性,对称性等(图 1)。

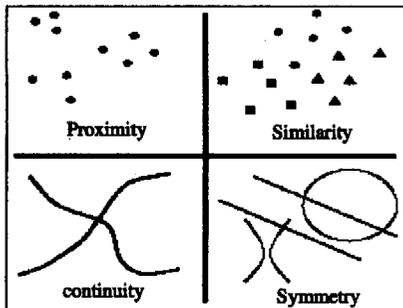


图 1 Gestalt 归类原则

人们从背景中分出物体时,轮廓线是一个重要的图形因素。轮廓线实际上不是线,而是人从一视点看物体时物体与背景相关部分在明暗、色彩、质感等方面的差别所造成的视感分界线(这种视感的产生与色彩知觉、深度知觉有关),依据轮廓线,人们可以简单地将物体与背景区分开来,并把复杂物体的结构及各部分关系看清楚。对于图形感知来说,这些线,特别是关键部位的线给人的信息最多,最能反映出物体的特征,是图形感知中最重要的部分。

在计算机视觉中,对感知归类的研究转化为物体识别中如何进行特征聚集的研究。受到心理学研究成果的激发,特别是 Gestalt 学派,在对物体进行识别前先将特征元素进行分组,通过分组提取图像中的轮廓线。心理学家们对感知组织进行学习,将人类视觉系统的分组现象进行分类,该分类依据 Gestalt 归类的规则。这些规则也可以作为将对对象特征进行分组的依据,这些特征包括边缘,拐角,以及区域等^[5]。

3 感知归类准则在图像特征提取中的应用

采用感知归类方法提取图像的轮廓线特征,首先需要提取轮廓最底层的支持元素——图像中的直线段。以直线段为初始级的构成元素,逐步提取下列分等级的特征,包括:直线段、较长线性线、相交线(包括“L”结点、“U”节点)、并行线,以及并行组等。在提取这些特征时,运用到感知归类中的接近性,连续性,相似性,并行性等规则。

3.1 提取图像中的直线段

Burns 直线检测器可以用来提取直线分段^[3]。为了消除零碎的直线对图像特征提取的影响,需要检测提取出的直线长度,只有满足以下公式的直线才能被保留。

^{*})该论文受重庆大学研究生创新基金“基于内容的图像检索引擎”项目的资助,编号 200504Y1A0070113。吴中福 教授,博士生导师,主要研究方向:计算机网络与通信、计算机网络安全。魏丽 硕士研究生,研究方向:基于内容的图像检索。李云 博士研究生,研究方向:模式识别。古毅 硕士研究生,研究方向:图像数据库检索,多维索引技术。

$$\epsilon(L_i) > \delta_i \quad (1)$$

其中 $\epsilon(\cdot)$ 代表了直线段 L_i 与图像中存在的最长线段的比率, 而 δ_i 是一个阈值。通过这种限制阈值的方式, 对提取出的直线做一定的筛选, 以保证提取出的特征具有一定的代表性, 消除一些噪声数据对图像处理过程的影响, 使处理集中于更具意义的代表性元素上。

3.2 利用邻近性提取较长线性线

在实际应用中, 由 Burns 算子提取的直线段是一些离散的线段, 应该将其分组以获得更高粒度的较长线性线。

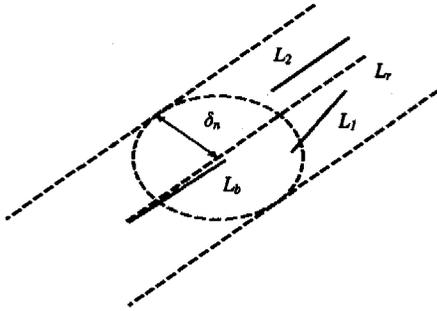


图2 较长线性线

在以 L_0 为中心轴的宽度为 $2\delta_n$ 的对称正交延伸区域 S_f 作为直线段的搜索区中进行提取较长线性线的操作, 区域中邻近的原始直线段被代表性直线 L_r 所代替, S_f 区域内的直线满足下面的条件:

$$(a) A(L_0, L_i) < \delta_n \quad (2)$$

$$\max\{D_0(L_0, e_{i_1}), D_0(L_0, e_{i_2})\} < \delta_n \quad (3)$$

$$(b) D(e_i, e_j) < \delta_n \quad (4)$$

$$\text{或 } \vartheta(\varphi_{L_i}(L_i), L_j) > 0 \quad (5)$$

其中 L_0 代表基础分段, L_i, L_j 代表了 S_f 中的任意两个分段, $A(\cdot)$ 代表了两个分段间弧度差绝对值的较小值, δ_n 是阈值。上面的公式中 e_i, e_j 分别是直线段 L_i 和 L_j 的两个端点, (e_{i_1} 和 e_{i_2} 是 L_i 的两个端点), $D_0(\cdot)$ 表示端点到线段的正交距离, $D(\cdot)$ 表示 L_i 和 L_j 相邻端点间的距离。式(5)中的 $\varphi_{L_i}(L_i)$ 是将 L_i 正交映射到 L_j 上, 而 $\vartheta(\cdot)$ 输出映射后两直线重叠的长度。

式(2)和式(3)保证了 S_f 中所有的线段与 L_0 近似在同一条直线上。式(4)和式(5)要求 S_f 中的所有线段必须满足相互邻近(端点在半径为 δ_n 的邻近圆形区域内), 或者至少与 S_f 中一条直线正交映射后重叠, 两条规则都是为了确保连续性。

为了确定直线 L_r , 我们需要一个 L_r 经过的点(找出 L_r 的中间点), 它的方向性, 以及它的长度。 L_r 的中间点和方向性分别是 S_f 中所有线段的加权平均值。权重是由线段长度决定的。为了获得 L_r 的长度, S_f 中所有线段的端点被正交投影到 L_r 上, 两个相隔最远的端点将作为 L_r 的端点。

此过程不断继续, 直到没有出现新的合并。只有确保搜索区域中只存在有限的线段才能保证此过程能够终止, 每一轮合并过程中线段数目不断减少直至循环终止。接着分析所有获得的直线, 只有满足下面公式的直线才能被保留下来:

$$\Gamma(L_i) < \delta_i \quad (6)$$

其中 $\Gamma(L_i)$ 是 L_i 的长度, 而 δ_i 是阈值。因为所获得的较长线性线由近似共线的直线段构成, 这些直线段间可能相互重叠, 也可能十分邻近。而获得较长线性线以后, 仍然需要进行剪枝操作, 消除一些较短对象。在此处设置阈值正是为了指导

剪枝的操作。

3.3 利用连续性提取连接点

连接点是指任意两个直线段的交汇处, 而此交汇处可能位于其中一条直线段上, 也可能没有在两条交汇直线的任何一条上。此处有一个要求, 两条交汇线之间的夹角必须在一定的范围内 $\theta_{\min} - \theta_{\max}$ 。以此避免出现两条基本平行或共线的直线。在文[8]中提到一些连接的类型。包括“L”连接, “U”连接等。“L”连接是由这些内部交角接近 $\frac{\pi}{2}$ 的共交点直线组成。这种连接是图像中突出拐角的有力证明。两个“L”连接的结合可形成一个“U”连接。“U”连接是由建筑物中规则的人造结构如窗口, 门等形成。有两种可能的“U”连接: (1)通过共线的L连接形成的“U”连接, (2)没有共线而形成的。对于第一种情况, 两个“L”连接中没有共线的直线应该具有相同的方向以保证形成有效的“U”连接。

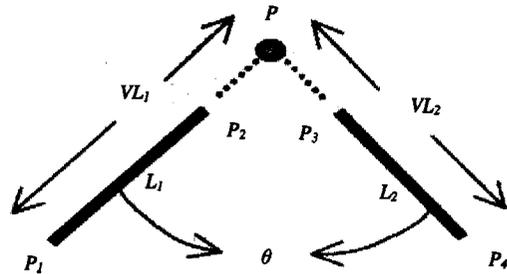


图3 两直线相交

现在对相交(或称两直线的相交)进行定义: 相交点 P 在两直线的端点处, 图3是一个典型的相交示意图。注意, 图中有两条虚拟的直线共享端点 P 。端点 P_1 和 P_4 分别是 P_2 和 P_3 的对立端点, 而 P_2 和 P_3 分别是虚拟相交线的另外两个端点。这样我们可以得出对于此连接程度的数学评价。

$$Q_i = \frac{|L_i| \cdot |\sigma_i^{\perp}|}{|VL_i|} \cdot \frac{|L_2| \cdot |\sigma_2^{\perp}|}{|VL_2|} \quad (7)$$

其中 $|VL_i|$ ($i=1, 2$) 代表虚拟直线的长度, 如图2所示。标准方差 σ_i^{\perp} 和 σ_i^{\parallel} 分别用于平衡两条直线端点位置的不确定性以及直线正交方向。这两个量可以由常数代替, 而不影响基本的分组算法^[9]。

所有其它的特征都是在较长线性线的基础上进行提取的。并行分组的提取是通过限制并行线 x, y 轴方向上的正交投影后叠交的数量而获得的, 同时考虑到局部方向性和并行组内部的方向性。“显著并行组”是对并行组做进一步的限制而获得, 通过调整视角的倾斜度, 显著并行组中至少应包括一条附属于L连接或U连接的直线。

4 特征向量的表示

总的来说, 通过感知归类提取图像特征后, 图像的特征向量可以按照如下的形式表示:

$$X = (x_1, x_2, x_3)^T \quad (8)$$

其中:

$$x_1 = \frac{\text{“L”连接中的直线总数}}{\text{构成各种形状的较长线性线的总数}} \quad (9)$$

$$x_2 = \frac{\text{“U”连接中的直线总数}}{\text{构成各种形状的较长线性线的总数}} \quad (10)$$

$$x_3 = \frac{\text{显著并行组中的直线总数}}{\text{构成各种形状的较长线性线的总数}} \quad (11)$$

对式(9), (10), (11)中分子被保留直线的总数进行标准

化,以此确保图像特征间公平的比较。在 $x_i, i \in \{1, 2, 3\}$ 中出现的共有直线只被计算一次。

显然 $x_i \in (0, 1)$, 其中 $i \in \{1, 2, 3\}$ 。此公式可看作将一幅图像映射到边界长度为 1 的立方体特征空间中。特征向量 X 代表了在此空间中图像的映射图像。

5 实验结果

图 4 为一幅原始的灰度图像,图 5 为利用感知归类中的方法提取原始图像边缘线的二维表示。由于阈值的选择可能存在着一定的误差,因此提取出的图像中仍然存在着一些较小的线段,将在下一步工作中继续对阈值的选择做进一步研究,希望能够找到一种方法,系统可以自动地调整阈值,而不是通过指定的阈值进行图像特征的提取。



图 4 原始灰度图像

结论 感知归类可以用于大量视觉信息以及图像分割技术中,它构成了计算机视觉处理的一个重要方面。本文着重研究了人类感知组织认知原理在图像边缘线性特征提取中的作用,利用元素间相似性、接近性、连续性、并行性等提取图像的显著特征。

本文未考虑到感知归类中一个非常重要的特征,即封闭性。多边形是由非并行线组成的封闭图像。一个多边形是一

个显著的图像关系。多边形是非偶然的图像关系,因为组成它们的边界线都是非偶然的。因此,图像中的多边形代表了图像的显著结构特征。根据感知归类的封闭规则,人类的视觉特性总是试图将曲线完善成为一个封闭的区域。如何将封闭性作为图像特征向量表示中的一维,以获得更准确的图像特征表示,是值得深入探讨的问题。

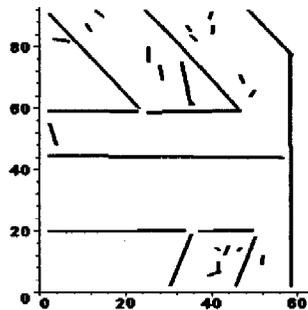


图 5 提取图像中边缘线

另外,由于建筑物一般都由拐角、窗户、门、边界等产生出大量的交叉点、平行线、平行群等,这些特征非常适合用感知归类的方法进行提取,下一步我们将实现在建筑物图像数据库中,利用感知归类的方法进行图像的检索和分类。

参考文献

- 1 完形心理学的视觉法则. [http://www. imageedu. com/Mode/Technic/Jiqiao/200502/1869. html](http://www.imageedu.com/Mode/Technic/Jiqiao/200502/1869.html)
- 2 王敏. 基于图象特征的物体形状识别方法研究. 华中科技大学控制科学与工程系. 机器视觉, 2004(8)

(上接第 221 页)

际中,我们可以假定当角色不知某个流的初始值时,认为该流是未定义的(undefined)。我们希望当前提条件是 undefined 时,规则 $A \Rightarrow B$ 的结果也是 undefined,也就是 A 为 undefined 时,推断出 B 也为 undefined。尽管这和经典的蕴涵关系不一样,但能简化实际的规则系统,也符合常规的推理习惯。我们定义一个特殊的模糊集合来表示 undefined,记作 θ :

$$\theta \triangleq \int_0/w \quad (12)$$

其中 W 表示区间, w 代表区间中的值。如果一个流 f 是 undefined,我们可以用 θ_f 来表示。

理论 1 A, A', B 以及 B' 分别是定义在 X, X, Y 以及 Y 上的模糊集,采用合成运算(compositional operation)“ $\vee - \wedge$ ”,如果:

$$A \Rightarrow B \quad (13)$$

$$A' = \int_0/x = \theta_A \quad (14)$$

那么:

$$B' = \int_0/y = \theta_B \quad (15)$$

证明:因为对于 X 中的任意 $x, A'(x) = 0$,

所以:

$$B'(y) = \sup_{x \in X} A'(x) \wedge I(A(x), B(y)) = 0 \quad (\forall y \in Y) \quad (16)$$

$$B' = \int_0/y = \theta_B \quad (17)$$

其中符号 \sup 表示取上限(upper bound), I 是模糊蕴涵算子

(fuzzy implication operator)。于是上面的金钱问题可以用式(18)来表示:

$$L_{\text{money}} = \theta_{\text{money}} \quad (18)$$

结论 在 John Funge 的区间取值认知流的基础上,本文介绍了一种新的应用于游戏以及 agent 认知建模上的认知流。该新认知流是在 John Funge 的区间上定义相应的语言变量,将该区间作为基础变量。通过使用语言变量,认知模型将会变得更简单。本文还给出了新认知流在解决不可数“可能世界”问题的方法,以及新认知流怎样用来推理不等式的方法。

参考文献

- 1 Funge J. Representing knowledge within the situation calculus using interval-valued epistemic fluents, J. Reliable Comput, 1999, 5 (1)
- 2 McCarthy J, Hayes P. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, eds. Machine Intelligence 4, Edinburgh University Press, Edinburgh, 1969. 463~502
- 3 Reiter R. The frame problem in the situation calculus, A simple solution (sometimes) and a completeness result for goal regression, In: Vladimir Lifschitz, ed. Artificial Intelligence and Mathematical Theory of Computation, Papers in Honour of John McCarthy, Academic Press, 1991. 359~380, 418~420
- 4 Lotfi A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning - II. Inf. Sci., 1975, 8(4): 301~357