

## 第二章

# 图和超图匹配的无监督学习



### 2.1 介绍

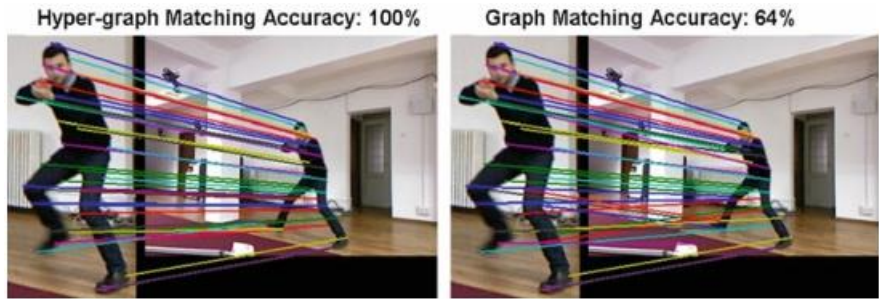
图和超图匹配是计算机视觉中的重要问题。它们被广泛应用于需要 2D 和 3D 特征匹配的各种任务，比如图像对齐、3D 重建、对象或动作识别等。图匹配控制了局部特征之间几何关系和外观的成对约束。同时，超图匹配包含了对特征集计算的高阶关系，可以同时获取几何信息和表层信息。因此，使用高阶约束可以使匹配问题更加准确、对非刚性变形和异常值的变化更加稳健。

时空域中的许多对象或其他实体可以用带有局部信息和全局信息的结点构成的图来表示。在许多计算机视觉任务中都需要寻找两种图形模型之间的对应关系，而且特征的类型在不同的应用程序中会有很大的不同。我们可以用图形匹配来识别形状，识别对象类型或者更复杂的问题，比如识别视频中的行为。图匹配的最终目的是在局部提取的特征和高阶边和超边计算的信息达成一致。事实上，结点的信息会从一个邻域传递到下一个邻域，最终保证全局对齐。匹配过程通常是迭代的，通过边或超边传递可能的结点的对应信息，最终将信息传播到两个图之间的整个所有节点分配集合中，最终达到结点到结点的匹配收敛。

近年来, 聚焦于图匹配的研究数量不断增加, 大多数新方法使用 IQP 公式或衍生公式。研究图匹配模型参数学习的论文并不多, 但是人们对这一问题的关注度正在提高。图匹配和高级约束成为超图匹配, 也收到越来越多的关注在计算机视觉领域, 这主要是由于更强大的几何建模可以获取更多的特征、提取到更高级的匹配。图匹配中的另一个非常相关的研究方向是如何在原始离散域中直接高效解决任务。在原始主题的论文中, 我们发现, 直接在正确的离散域进行优化通常对大于优化在宽松域中进行的全局优化, 之后还需要做一个简单的二值化。图匹配的另一个有趣的方向是寻找多个图之间的节点对应, 这将支持跨多个图像或对象模型的图匹配。多图任务在文献中也发现了一个值得注意的方法, 包括初始张量公式。

本章呈现的材料是基于之前的工作。由一组用于图和超图匹配的学习的方法组成, 通常认为在图和超图匹配的无监督和半监督学习方法是计算机视觉文献中最早提出的算法。

在本章中, 我们提出了图和超图匹配、优化、学习的方法。我们的学习和优化算法是基于相同的核心思想, 高阶公式是二阶图匹配情况的自然拓展。首先, 我们给出了图匹配的一般 QAP 公式, 然后将其扩展到超图匹配。本文提出了两种不同的图像匹配算法, 即光谱匹配算法和整数投影算法。



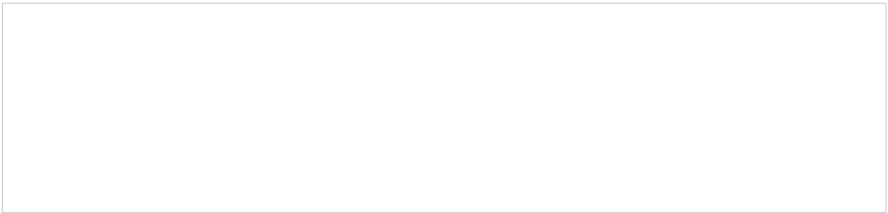
**Fig. 2.1** 超图匹配比图匹配功能更强大。在尺度发生剧烈变化的情况下，采用高阶约束匹配的集合和外观尺度不变模型具有更好的性能。

不动点法用于条件随机域中的图匹配和映射推理。这两种方法结合起来效果很好，第一种方法提供了松弛问题的最优解，第二种方法提供了一个强大的离散过程，具有重要的理论保证。同时，我们提出了一种新的超图匹配的监督学习和无监督学习方法。

我们对具有双约束的图匹配问题的学习和优化算法提供了理论和实践基础。我们提出了拓展到更高阶的方法，并提供了实验，显示了在合成图像和真实图像上的性能。本章统一介绍了图和超图匹配的学习和优化，并将在第三章中推广到图和超图的聚类。

### 2.1.1 与无监督学习原理的关系

上一章介绍的无监督学习的定理 5 和图形匹配联系最为密切。



这一原则对于图匹配情况下的推理和学习非常重要。在该任务中，对齐通常是指两个图的节点和边（或超边）之间在几何或外观上的一致性。它们也可能涉及到更高层次的语义抽象的相似性。然而，这种原则不经常发生。联系非常稀少，在空间和时间上没有形成强有力的协议集群。例如，一个对象的某些部分可能意外地与另一个无关对象的某些部分相似。在不断移动的世界中，随机模式可能类似于一些众所周知的物体（例如，云的结构可能类似于小狗）。然而，在空间和时间上稳定的两个结构是很难出现的。在形状、外观和完整的结构层面，它们通常表明在两个实体之间有一个有意义的联系。这种关系可以是在同一层面（例如，从不同的视角看，或者在两个不同的时间里看到的是同一个对象），或者在语义范畴的层面（例如，对象是同一种类的，具有相似的意义）。

我们将在本章中看到，节点和边的对齐在两个图的节点之间的候选分配的特殊图中建立了强链接。通过这样的强一致性链接，最有可能是正确的候选任务会形成一个强大的集群。因此，找到正确的赋值就变成了一个聚类问题，该问题通常需要遵守一定的匹配约束，这样一个图中的一个节点最多只能匹配另一个图中的一个节点。因此，图匹配和聚类之间的联系是自然的，这就是为什么为一个问题设计的算法（例如，谱聚类或用于图匹配的整数投影不动点）可以很容易地适应另一个问题（通常是约束子解向量）。

接下来，我们给出了图形匹配的数学公式。在此基础上，介绍了谱图匹配和整数投影不动点算法。在第三章中，它们将进一步扩展到高阶匹配情形（超图匹配），并适应于高阶约束的聚类任务。在这些章节中介绍的所有推理或学习方法都与聚类的一般概念有着根本的联系，聚类是所有无监督学习的核心。

## 2.2 图匹配

带有成对约束的图匹配问题包括求解指示器向量  $\mathbf{x}^*$ , 使它最大化具有特定映射约束的二次分数函数:

$$\mathbf{x}^* = \arg \max (\mathbf{x}^T \mathbf{M} \mathbf{x}) \text{ s.t. } \mathbf{A} \mathbf{x} = \mathbf{1}, \mathbf{x} \in \{0, 1\}_n, \quad (2.1)$$

$\mathbf{x}$  为指示向量,  $x_{ia} = 1$  如果特征  $i$  和另一幅图像的特征  $a$  匹配, 否则为 0;  $\mathbf{A} \mathbf{x} = \mathbf{1}, \mathbf{x} \in \{0, 1\}_n$  对  $\mathbf{x}$  施加一对一的约束, 这样一幅图像中的一个特征最多可以与另一张图像的另一个特征相匹配。M 为一个包含正元素的矩阵, 其中包含两两计函数使  $M_{ia;jb}$  度量一对特征  $(i, j)$  与另一幅图像的一对候选匹配  $(a, b)$  在几何和外观方面的一致性。

虽然光谱匹配算法 (SM) 要求 M 具有非负元素, 但整数投影不动点法没有这一要求。还要注意, 候选分配的局部一元项可以存储在 M 的对角线上; 在实践中, 我们注意到考虑到这样的局部特征, 一元信息在两两配对的分数  $M_{ia;ib}$  和在对角线上留下零产生了好的结果; 组合线性和二次项的略微不同的形式的匹配目标也是可能的:  $\mathbf{d} \mathbf{x} + \mathbf{x}^T \mathbf{M} \mathbf{x}$ 。这里, 向量  $\mathbf{d}$  的元素表示一元分数, M 包含两两约束。注意, 该公式不适用于光谱图匹配 (第 2.4.1 节), 但对 IPFP (第 2.4.2 节) 有效, 类似于成对马尔科夫网络中 MAP (Maximum A Posteriori) 推理问题的二次公式 [37-39]。主要区别在于, 对于映射推理任务, 解决方案上的约束  $\mathbf{A} \mathbf{x} = \mathbf{1}, \mathbf{x} \in \{0, 1\}$  通常是多对一的: 图中的多个节点可以具有相同的标签。许多算法, 包括我们的 IPFP, 可以处理这两个问题, 只需稍加修改。在这一章中, 我们提出了类似的学习算法, 适用于这两个问题, 图匹配和图推理, 以再次强调这两个任务是密切相关的, 并给出一个一般的, 包括我们的优化和学习方法。

回到主要的 QAP 图匹配公式, 我们认为  $M_{ia;jb}$  是本质上是一个由某个参数向量  $\mathbf{w}$  定义的函数。成对得分的  $M_{ia;jb}$  也会超图匹配密切相关。

$$M_{ia;jb} = \exp(-\mathbf{w}^T \mathbf{g}_{ia;jb}), \quad (2.2)$$

其中  $w$  是学习的参数权重的向量,  $gia; jb$  是一个通常包含非负误差和变形的向量, 用于描述当我们将特征  $(I, j)$  与特征  $(a, b)$  匹配时集合和外观的变化。

## 2.3 超图匹配

超图匹配通过将公式 2.1 扩展到能够包含高阶约束的张量公式, 自然地从业图匹配中得到。计算机视觉中最早的超图匹配方法是概率方法[24]和基于张量的高阶幂方法[22]。我们提出的公式适用于任何阶超图模型。然而, 在实践中, 三阶约束在效率和捕捉对象外观和几何图形的能力之间提供了一个很好的折衷方案。经典的图匹配方法处理在相似、仿射或射影变换下非不变特征之间的成对关系。三阶关系仍然可以有效地处理, 同时能够对高阶几何和外观进行建模。物体可以用三角网格来描述, 在三角形的内部计算出角点的特征和外观。在本章的其余部分中, 我们只讨论三阶超图匹配, 同样的公式在原则上可以扩展到任何阶关系的匹配。

给定两组特征, 一组来自模型图像  $I_m$ , 另一组来自测试图像  $I_t$ , 三阶(超图)匹配包括寻找这两个集合之间的对应关系, 从而使作为候选匹配三元组之和的匹配分数最大化。分数可以考虑潜在赋值的三元组上的几何和外观信息, 扩展用于图形匹配的成对分数。

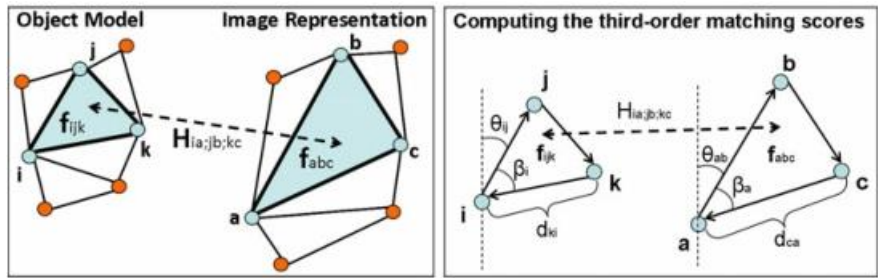
设置一个候选赋值列表  $L_{ia} = (i, a)$ , 每个  $(i, a)$  对应都有一个唯一的索引  $ia$ , 我们用一个指示符向量  $x$  (长度与  $y$  相同) 来表示一个可能的解, 这样如果模型中特征  $i$  与测试图像中的特征  $a$  匹配, 则标记  $x_{ia}=1$ , 否则, 为 0。通常还会施加一对一的映射约束, 这样一个图像中的一个特征可以匹配另一个图像中的单个特征。映射约束以矩阵的形式表示

$Ax \leq 1, x \in \{0, 1\}_n$ ,  $A$  为一个二值矩阵。三级匹配分数为:

$$S(x) = ia; jb; kc H_{ia; jb; kc} x_{ia} x_{jb} x_{kc}. \quad (2.3)$$

在这里,  $H$  是一个超对称传感器, 具有非负向性. 它们随着特征元组之间匹配的质量而增加,  $H_{iaJB; kc}$  表示模型图像的特征  $(I, j, k)$  在外观和质量方面的匹配程度。

图 2.2 左:使用三阶表示(同时考虑几何和外观, 根据三角形计算匹配得分)将模型与图像进行匹配。右:



三阶分数  $H_{ia;jb;kc}$  可以是几何和外观的函数。几何信息可以用旋转不变量角、旋转不变量角和成对距离来表示, 通过除以周长则成对距离可以保持比例不变。外观可以用特征向量  $f_{ijk}$ 、编码梯度、颜色或纹理信息来表示。每个线索的相对重要性是依赖于应用的, 应该是在训练过程中生成的。利用特征  $(a, b, c)$  求解超图匹配意味着在一对一映射的约束下, 找到最大化  $S(x)$  的解。

$$x^* = \arg \max_{ia;jb;kc} H_{ia;jb;kc} x_{ia} x_{jb} x_{kc}. \tag{2.4}$$

在实践过程中, 人们可以使用任何类型的信息来设计  $H_{ia}$  分数。大多数人更多关注的是几何的信息, 但是在这里我们通过学习更强大的分数, 同时考虑几何和外观上的分数, 可以显著提高匹配性能, 我们使用的分数是:

$$H_{ia;jb;kc} = \exp(-w^T g_{ia;jb;kc}), \tag{2.5}$$

其中  $w$  是要学习的权重/参数向量,  $g_{ia;jb;kc}$  是一个包含非负误差/变形的向量, 当模型图像  $I_m$  中的特征  $(I, j, k)$  定义的三角形与测试图像  $I_t$  中的三角形  $(a, b, c)$  匹配时, 对几何形状和外观的变化进行建模。如图 2. 2.

### 2.4 求解图匹配

图与超图匹配是 NP 困难问题, 必须寻找有效的匹配方法近似方法。这里, 可以运用 WebrieFlyReviewUmain 方法进行图形匹配、光谱匹配[16]和整数投影不动点 (IPFP) [6], 这些方法与在超图上的工作有关, 这些工作将在本章后面部分讨论。

### 2.4.1 光谱匹配

自其发表以来,我们的光谱图匹配方法已成功应用于广泛的计算机视觉应用,对研究工作和商业专利都产生了重大影响。这里,我们列举了光谱图匹配在各个科学领域中的一些有代表性的研究方向:视频中物体交互作用的动作识别与建模[40-44];物体类别的发现与建模[45-50];捕捉三维人体行为[51];三维场景采集[52];二维匹配物体[53],对称性检测与分析[54, 55],三维形状集合中的自动注释[56],医学图像中的沟脑皮层模式检测[57-59],医学图像注册[60],多蛋白质相互作用网络的对齐[61],基于WiFi的室内定位[62],人的重新识别[63],监控视频总结[64],检测和跟踪皮损[65]等。

光谱匹配也是发展中的谱匹配算法的起点,如仿射约束谱匹配(SMAC)[3],整数投影不动点(IPFP)[6],张量高阶匹配[22],概率和近似谱图以及超图匹配[8, 24, 66-68],以及基于谱松弛的地图推理算法[37, 38]。

光谱匹配优化的解决了问题 2.1 中的如下松弛变量。

$$\mathbf{x}^* = \arg \max (\mathbf{x}^T \mathbf{M} \mathbf{x}) \quad \text{s.t. } \mathbf{x}^T \mathbf{x} = 1. \quad (2.6)$$

该问题的解由  $\mathbf{M}$  的第一特征向量给出,由于  $\mathbf{M}$  只有正元素,根据 Perron-Frobenius 定理,特征向量元素也是正的,这使得特征向量的后处理离散化变得容易。我们观察到  $\mathbf{M}$  可以解释为一个图的邻接矩阵,它的节点代表  $M_{ia}; j_b$  边在这些可能的赋值之间存在一致性。这个图有一个特殊的结构,这有助于我们理解为什么使用第一个特征向量近似解来找到问题 2.1 的近似解它包含以下内容:

1.

强连通簇主要是由正确的分配形成,这些分配倾向于在彼此之间建立协议链接(强边)。当成对的分配在对应的特征之间的成对关系(例如几何关系)上达成一致时,就形成了这些一致链接(见图 2.3)。

2.

很多不正确的赋值,大部分是在簇外或与簇弱连接(通过弱边),由于它们建立协议链路的概率很小,并且以随机、非结构化的方式形成这些链路,因此不能形成强连接簇。



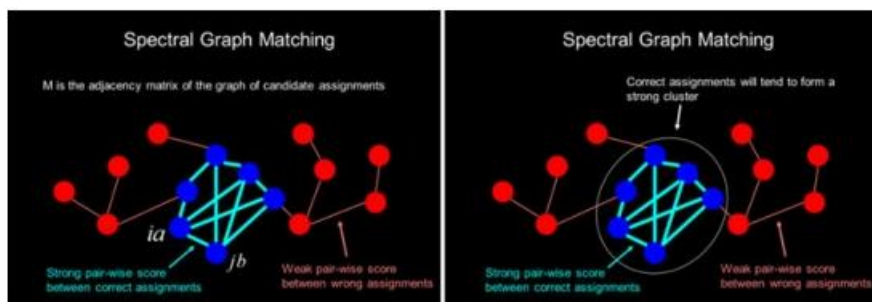


Fig. 2.3 左图：例如，图中的一个节点可能来自一个匹配的节点。将图匹配定义为整数二次规划（式 2.1）的矩阵  $M$  成为候选赋值图的邻接矩阵。右：更有可能建立牢固的协议链接（图中的蓝色粗线）。因此，一个自然的聚类是由图中正确的候选赋值组成的，这些候选赋值可以通过谱方法快速找到，因为它们对应于  $M$  的主特征向量中的大的值。

我们可以使用各种离散化程序来寻找近似解。一种思想是应用匈牙利方法，它可以有效地找到满足映射约束的二元解，并使特征向量为  $M$  的点积最大化。另一种想法是使用我们最初在 Leordeanu 和 Hebert[16]中提出的贪婪离散化算法：我们解释主元的每个元素  $M$  的特征向量  $v$  作为相应赋值正确的置信度。我们选择最大置信度的元素作为开始，然后我们删除所有冲突的赋值，并选择正确的赋值，然后重复这个过程，直到所有赋值都被标记为正确或不正确。图匹配问题的特征向量松弛，结合贪婪离散化过程，使得谱匹配成为最有效的成对约束匹配算法之一。

### 2.4.2 整数投影不动点法

在最近的一篇论文中，Leordeanu 等人提出了另一种有效的图匹配算法 (IPFP)，它优于许多最先进的方法。自出版以来，IPFP 的修改版本已经应用于任务，如分割[69]，更高层次的视频表示[70]，超序磁共振成像[71]，艺术图像分析[72]，概括 IPFP 算法的离群点检测[73]，以及大规模图形匹配的 FastIPFP[74]，Zaslavskiy 等人独立开发了一种算法，该算法在不同的公式中应用了 IPFP 的许多性质。

IPFP 可以作为一个独立的算法使用，也可以作为其他图匹配算法的离散化过程，例如谱匹配。此外，IPFP 可用于 MRF 中的映射推理，表述为二次整数规划问题。它有效地解决了(虽然不一定是最佳的)问题 2.1 的一个更严格的松弛问题，也是众所周知的 NP 难问题：

$$\mathbf{x}^* = \arg \max (\mathbf{x}^T \mathbf{M} \mathbf{x}) \quad \text{s.t. } \mathbf{A} \mathbf{x} = \mathbf{1}, \mathbf{x} \geq \mathbf{0}. \quad (2.7)$$

公式 2.1 和 2.7 之间的唯一区别是后者中，允许取连续值。整数投影不动点法（IPEP）将任何初始的解（连续或者离散）作为输入，可以快速找到问题 2.1 的初始离散约束的解，该解比初始解的得分通常显著更高。

## 算法 2. 高阶匹配的循序二阶展开

---

### Algorithm 2.1 Sequential second-order expansion for higher order matching

---

Initialize  $\mathbf{x}^* = \mathbf{x}_0$ ,  $S^* = \mathbf{x}_0^T \mathbf{M} \mathbf{x}_0$ ,  $k = 0$ , where  $x_i \geq 0$  and  $\mathbf{x} \neq \mathbf{0}$ ;  
**repeat**  
  1) Let  $\mathbf{b}_{k+1} = P_d(\mathbf{M} \mathbf{x}_k)$ ,  $C = \mathbf{x}_k^T \mathbf{M} (\mathbf{b}_{k+1} - \mathbf{x}_k)$ ,  $D = (\mathbf{b}_{k+1} - \mathbf{x}_k)^T \mathbf{M} (\mathbf{b}_{k+1} - \mathbf{x}_k)$ ;  
  2) If  $D \geq 0$ , set  $\mathbf{x}_{k+1} = \mathbf{b}_{k+1}$ . Else let  $r = \min(-C/D, 1)$  and set  $\mathbf{x}_{k+1} = \mathbf{x}_k + r(\mathbf{b}_{k+1} - \mathbf{x}_k)$ ;  
  3) If  $\mathbf{b}_{k+1}^T \mathbf{M} \mathbf{b}_{k+1} \geq S^*$  then set  $S^* = \mathbf{b}_{k+1}^T \mathbf{M} \mathbf{b}_{k+1}$  and  $\mathbf{x}^* = \mathbf{b}_{k+1}$ ;  
  4) Set  $k = k + 1$   
**until**  $\mathbf{x}_{k-1} = \mathbf{x}_k$   
**return**  $\mathbf{x}^*$

---

在这里，步骤 2 中的  $P_d(\cdot)$  表示对离散域的投影，下面将讨论。

---

## 2.5 理论分析

**命题 2.1** 对于任意的向量  $\mathbf{x}$  属于  $\mathbf{R}^n$ ，在有二元元素的问题 2.2 的定义域中存在  $\mathbf{x}^T \mathbf{M} \mathbf{y}$  的全局最优  $\mathbf{y}$  (因而也在问题 2.1 的定义域中)。

**命题 2.2** 二次分数  $\mathbf{x}^T \mathbf{M} \mathbf{x}$  在每一步  $k$  出增加， $\mathbf{x}_k$  的序列收敛。

**命题 2.3** 如果  $\mathbf{M}$  是带正元素的正半定，那么算法在有限次迭代中收敛到一个离散解，这是问题 2.2 的最大值。

**命题 2.4** 如果  $\mathbf{M}$  有非负元素且秩为-1，那么算法在第一次迭代后会收敛并返回原问题的全局最优。

## 2.6 求解超图匹配

超图匹配和图匹配一样，都是 NP 难的问题，因此人们通常需要高效的算法来寻找好的近似解。在从二阶匹配到高阶匹配的过渡过程中，复杂度急剧增加。因此，开发快速准确的算法变得越来越困难。一些当前的方法[22-24]利用秩一近似（ROA）来降低复杂性。这导致要么放松原始问题的一对一匹配约束[22]，要么根据概率解释修改公式[23,24]。

在现实世界的计算机视觉应用中，关于高阶匹配的论文屈指可数。一个例子是在医学图像中发现皮肤损伤，即使在图匹配的情况下，矩阵  $M$  也可能很大。只有精心设计、高效的方法，利用匹配问题的内部统计信息，才能通过构造和处理稀疏矩阵或在算法上降低问题的复杂性来处理这项任务。这可以通过分解矩阵[9]或利用某些局部几何变换约束的代数操作来实现[81]。在研究 NP 难超图匹配问题时，必须记住，无论是在存储还是计算方面，任务都是非常昂贵的。

在本节中，我们提出了一种新的超图匹配方法，该方法可以独立使用，从统一的初始解开始，也可以作为一个细化过程，从另一个算法返回的解开始。与其他方法相反，我们的目标是最大化原始目标得分，同时尽可能保留原始离散一对一匹配约束。我们通过实验证明，原始区域的局部优化比松弛区域的全局优化更有效，Leordeanu 等人也在二阶匹配的情况下观察到了这一事实。

我们提出的方法（算法 2.2）是一个迭代过程，在每个步骤  $k$ ，通过其围绕当前解  $x_k$  的二阶泰勒展开来近似高阶分数  $S(X)$ 。这将超图匹配问题转化为一个整数二次规划（IQP），它是围绕当前解决方案  $x_k$  定义的。二阶近似也可以局部优化，在连续域  $Ax = 1, x \in [0, 1]^n$  中相当有效。

给定连续域  $Ax = 1, x \in [0, 1]^n$  中的一个可能解  $x$ ，设矩阵  $M(x)$  是  $x$  的函数，通过如下边缘化张量  $H$  得到：

$$M(\mathbf{x})_{ia; jb} = \sum_{kc} H_{ia; jb; kc} x_{kc}. \quad (2.8)$$

同样，我们定义列向量  $\mathbf{d}$ :

$$d(\mathbf{x})_{ia} = \sum_{jb; kc} H_{ia; jb; kc} x_{jb} x_{kc}. \quad (2.9)$$

有了定义的  $M$  和  $d$ ，我们可以按如下方式写出第二阶泰勒近似第三阶匹配分数(公式 2.5)：

$$S(\mathbf{x}) \approx S(\mathbf{x}_0) + 3(\mathbf{x}_T^T \mathbf{M}(\mathbf{x}_0) \mathbf{x} - \mathbf{d}(\mathbf{x}_0)^T \mathbf{x}). \quad (2.10)$$

我们可以通过优化二次赋值  $\mathbf{x}_0(\mathbf{x}_0)^T \mathbf{x} - \mathbf{d}(\mathbf{x}_0)^T \mathbf{x}$  在连续域  $\mathbf{ax} = 1, \mathbf{x} \in [0, 1]^n$  来局部提高  $\mathbf{x}_0$  附近的分数，因为第一项  $S(\mathbf{x}_0)$  是常数。注意，实际上矩阵  $M$  往往接近半正定。在这种情况下，二次近似是一个凹极小化问题，通常很难全局求解(与凸极小化相反)。因为在每次迭代中，局部最大化就是 Algorithm 2.2).

### Algorithm 2.2 高阶匹配的顺序二阶展开

---

#### Algorithm 2.2 Sequential second-order expansion for higher order matching

---

```

 $\mathbf{x} \leftarrow \mathbf{x}_0$ 
 $\mathbf{x}^* \leftarrow \mathbf{x}_0$ 
 $S^* \leftarrow \sum_{ia,jb,kc} H_{ia,jb,kc} x_{ia} x_{jb} x_{kc}$ 
repeat
   $M_{ia,jb} \leftarrow \sum_{kc} H_{ia,jb,kc} x_{kc}$ 
   $d_{ia} \leftarrow \sum_{jb,kc} H_{ia,jb,kc} x_{jb} x_{kc}$ 
  repeat
     $\mathbf{b} \leftarrow \pi(2\mathbf{M}\mathbf{x} - \mathbf{d}^T \mathbf{x})$ 
     $S \leftarrow \sum_{ia,jb,kc} H_{ia,jb,kc} b_{ia} b_{jb} b_{kc}$ 
    if  $S > S^*$  then
       $S^* \leftarrow S$ 
       $\mathbf{x}^* \leftarrow \mathbf{b}$ 
    end if
     $t^* = \arg \max_{t \in [0, 1]} \mathbf{x}(t)^T \mathbf{M} \mathbf{x}(t) + \mathbf{d}^T \mathbf{x}(t), \quad t \in [0, 1]$ 
    where  $\mathbf{x}(t) = \mathbf{x} + t(\mathbf{b} - \mathbf{x})$ 
     $\mathbf{x} \leftarrow \mathbf{x} + t^*(\mathbf{b} - \mathbf{x})$ 
  until convergence
until  $S^*$  does not improve
return  $\mathbf{x}^*$ 

```

---

通过跟踪最佳离散解  $\mathbf{x} \leftarrow \mathbf{b}$ ，我们的方法总是返回至少与起始解一样好的解。由于分数  $S$  在上面有界，离散解的个数是有限的，保证算法提高匹配分数，有限次迭代后停止。如果内部循环的上一次迭代没有改进，算法将停止。在我们的实验中，收敛平均需要少于五次的迭代。

## 2.7 学习图匹配

在本节中，我们介绍了图匹配方法的监督、半监督和无监督学习变体以及条件随机场的详细参数学习。

### 2.7.1 理论分析

我们提出的学习算法是由矩阵  $M$  和它的主特征向量  $v$  的统计特征驱动的，主特征向量  $v$  是由谱匹配算法给出的连续解。为了从理论上分析性质，我们需要一些假设和近似值。我们做得假设是直观的。匹配问题的每一个实例都是唯一的，所以绝对不能肯定地说  $M$  和它的特征向量  $v$ ，也不能说返回的解唯一。因此，我们关心的是  $M$  的平均性质，而不是特殊情况。对此，我们提出了一个模型，并通过实验进行验证。

对于给定的匹配实验及其对应的矩阵  $M$ ，设  $p_1 > 0$  为正确的情况。同时，设  $p_0 = E > 0$ ，如果至少其中一个分配是可变的。 $p_1$  应该比  $p_0$  高，因为正确的任务应该在外观和几何上一致，而且有很高的二阶分数。而错误的任务只有在偶然的情况下才会出现高的分数。直觉上，我们希望  $p_1$  越高  $p_0$  越低，匹配度越高。我们还希望性能取决于它们的比值  $pr = p_0/p_1$ ，而不是它们的绝对值，因为将  $M$  乘以常数不会改变前导特征向量。同样，我们定义了

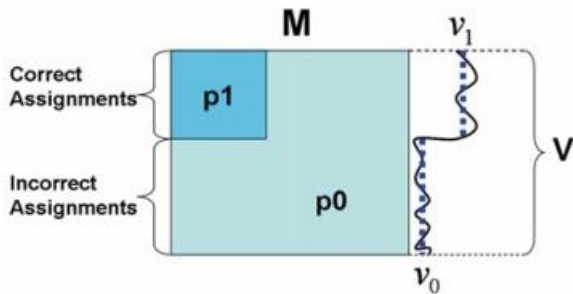


Fig. 2.4 正确赋值之间的成对得分(匹配矩阵  $M$  的元素)比至少有一个错误赋值的元素具有更高的期望值  $p1$ , 平均值为  $p0$ . 这将反映在特征向量  $v$  中, 该特征向量将具有更高的平均值  $v1$  用于校正

平均特征向量值  $v1 = E(v_{ia})$  过度矫正分配  $ia$ , 和  $v0 = E(v_{jb})$ , 过度错误分配  $jb$ . 频谱匹配算法假设正确分配对应于特征向量  $v$  的较大的值, 错误分配对应于  $v$  的低值, 因此  $v1$  越高,  $v0$  越低, 匹配率越好. 在  $pr$  的情况下, 如果我们能够在学习期间最小化训练序列中所有图像对的平均比率  $v_r = v0/v1$  (因为特征向量的范数无关), 那么我们将期望优化整体训练匹配率. 该模型假设完全连通图, 但是可以验证我们接下来获得的结果对于弱连通图也是有效的, 这也在我们的实验中示出.

对于给定的图像对, 需要研究  $v_r$  和  $pr$  之间的关系. 我们知道  $\lambda v_{ia} = \sum_{jb} M_{ia,jb} M_{jb,v}$ . 为了表达清楚, 我们假设对于左图像中的  $n$  个特征中的每一个, 在右图像中有  $k$  个候选对应. 我们做如下估计:  $E(\sum_{jb} M_{ia,jb} v_{jb}) \approx \sum_{jb} E(M_{ia,jb}) E(v_{jb})$ , 考虑到任何  $v_{jb}$  几乎独立于任何特定的  $M_{ia,jb}$ , 因为  $M$  很大. 近似值实际上是一个“ $\geq$ ”不等式, 因为相关性预计为正(但非常小). 对于我们给定的矩阵  $M$ , 我们称它的第一特征值为  $\lambda$ . 因此, 对于正确的对应关系  $ia$ ,  $\lambda E(v_{ia}) = \lambda v1 \approx np1v1 + n(k-1)p0v0$ . 同样, 如果  $ia$  是错误的对应关系, 那么  $\lambda E(v_{ia}) = \lambda v0 \approx np0v1 + n(k-1)p0v0$  将两个方程除以  $p1v1$ , 取两者之比, 我们得到

$$v_r \approx \frac{pr + (k-1)pr v_r}{1 + (k-1)pr v_r} \quad (2.11)$$

求解这个二次方程, 我们得到

$$v_r \approx \frac{(k-1)pr - 1 + \sqrt{(1 - (k-1)pr)^2 + 4(k-1)pr^2}}{2(k-1)pr} \quad (2.12)$$

使用等式 2.11 和 2.12, 可以验证对于  $k > 1$ ,  $v_r$  是  $pr$  的单调递增函数. 这实际上并不令人惊讶, 因为我们预期  $pr = p0/p1$  越小,  $v_r = v0/v1$  越小, 特征向量  $v$  将越二进制定化(并且更接近二进制基本事实  $t$ ), 错误赋值的元素接近 0. 正如我们在图 2.6、2.7 和 2.8 中的实验所示, 这种近似在实践中非常精确. 此外, 虚拟现实越小, 期望匹配率越高, 也就是说, 正确匹配的特征数除以特

征总数。为了清楚起见，在分析过程中，我们假设两幅图像中的特征数量相等。我们还假设左图像中的每个特征在右图像中都有一个正确的匹配。然而，正如我们在实验部分所展示的，我们的算法对于异常值的存在是鲁棒的。

最小化  $\mathbf{v}r$  的一种方法是最大化  $\mathbf{v}$  和地面真实指示向量  $\mathbf{t}$  之间的相关性，同时确保左图像中的一个特征匹配左图像。然而，在第二章中，我们希望以一种无人监督的方式最小化虚拟现实，这种方式在训练过程中不知道  $\mathbf{t}$ 。我们建议的解决方案是最大化  $\mathbf{v}$  和它的二进制版本之间的相关性。

## 2.7.2 图匹配的监督学习

我们希望找到几何和外观参数  $\mathbf{w}$ ，其最大化(在有监督的情况下) $M$  的主特征向量和地面真实  $\mathbf{t}$  之间的预期相关性，该相关性在经验上与所有训练图像对上的以下总和成比例：

$$J(\mathbf{w}) = \sum_{i=1}^N \mathbf{v}^{(i)}(\mathbf{w})^T \mathbf{t}^{(i)}, \quad (2.14)$$

其中  $\mathbf{t}^{(i)}$  是第  $i$  个训练图像和真实值相比的指示向量。我们通过坐标梯度上升最大化  $J(\mathbf{w})$ ：

$$\mathbf{w}_j^{(k+1)} = \mathbf{w}_j^{(k)} + \eta \sum_{i=1}^N \mathbf{t}_i^T \frac{\partial \mathbf{v}_i^{(k)}(\mathbf{w})}{\partial \mathbf{w}_j}. \quad (2.15)$$

为了简化本章其余部分的符号，我们使用了  $F$  表示任何向量或矩阵  $F$  相对于  $\mathbf{w}$  的某个元素的导数的向量或矩阵。在下一节中会给一种求对称矩阵特征向量偏导数的可能方法(当  $\lambda$  为 1 阶时)。8.8 马格努斯和诺德克尔[83]和库尔等人在谱聚类的背景下提出了：

$$\mathbf{v}' = (\lambda \mathbf{I} - \mathbf{M})^\dagger (\lambda' \mathbf{I} - \mathbf{M}') \mathbf{v}, \quad (2.16)$$

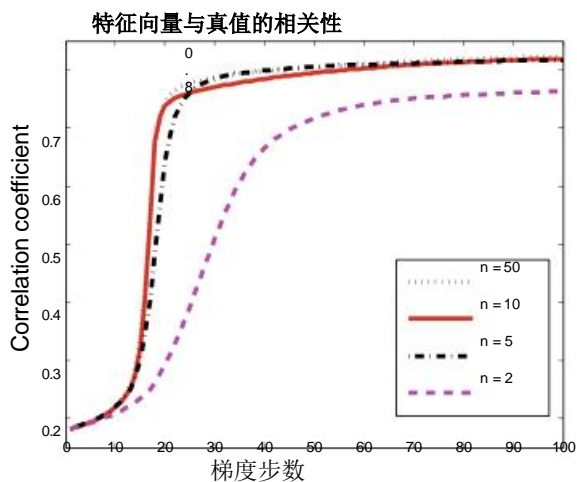
其中  $\mathbf{A}^\dagger$  表示  $\mathbf{A}$  的伪逆，并且：

$$\lambda' = \frac{\mathbf{v}^T \mathbf{M}' \mathbf{v}}{\mathbf{v}^T \mathbf{v}}. \quad (2.17)$$

这些方程是利用  $M$  是对称的以及  $\mathbf{v}^T \mathbf{v}' = 0$  和  $M\mathbf{v} = \lambda \mathbf{v}$  的等式得到的。然而，这种方法是一般的，因此没有充分利用在这种情况下  $\mathbf{v}$  是一个具有大特征集的矩阵的主特征向量这一事实。由于  $M - \lambda \mathbf{I}$  算法规模大且秩不足，在实际应用中对其伪逆的计算效率不高。相反，我们使用幂方法来计算近似主特征向量的偏导数： $\mathbf{v}' = \mathbf{v} M$ 。这与 Bach and Jordan 有关[851]，但在 Bach and Jordan[851]中，该方法用于分割，Cour 等人也指出[841]。在这种情况下，它可能非常不稳定，因为在分割和典型聚类问题中，前两个特征值之间的特征集并不大。

**Fig. 2.5**

该图显示了用于计算特征向量的近似导数的不同递归迭代次数  $n$  的特征向量和地面真实解之间的归一化相关性(70 次实验的平均值)。即使  $n$  小至 5, 学习方法也以相同的方式收敛, 返回相同的结果。





### 2.7.3 用于图匹配的无监督和半监督学习

无监督学习的思想(在第 2.7.1 节中介绍)是最大化  $\mathbf{v}^*$  而不是  $\mathbf{v}_r$ , 这可以通过最大化特征向量和从特征向量获得的二进制解之间的点积来实现。因此, 在无监督训练期间, 我们最大化以下式子:

$$J(\mathbf{w}) = \sum_{i=1}^N \mathbf{v}^{(i)}(\mathbf{w})^T \mathbf{b}(\mathbf{v}^{(i)}(\mathbf{w})). \quad (2.20)$$

这里的困难是  $\mathbf{b}(\mathbf{v}^{(i)}(\mathbf{w}))$  不是一个连续的函数, 也不可能用  $\mathbf{w}$  来表示, 因为  $\mathbf{b}(\mathbf{v}^{(i)}(\mathbf{w}))$  是一个迭代离散化过程的结果。然而, 重要的是  $\mathbf{b}(\mathbf{v}^{(i)}(\mathbf{w}))$  是分段常数, 并且除了有限组不连续点之外, 在任何地方都为零导数。因此, 我们可以预期, 我们将只在  $\mathbf{b}$  为常数且导数为零的点上计算梯度。此外, 在这些点上, 梯度步长将降低  $\mathbf{v}_r$  (等式 2.13), 因为  $\mathbf{b}$  中的变化(当梯度更新通过  $\mathbf{b}$  中的不连续点时)不会影响  $\mathbf{v}_r$ 。降低  $\mathbf{v}_r$  会增加  $\mathbf{v}_t$ , 也会降低  $\mathbf{m}$ , 因此无需担心  $\mathbf{b}$  的不连续点, 就能达到预期目标。这一点在我们的每次实验中都得到了验证。然后, 学习步骤函数变为:

$$\mathbf{w}_j^{(k+1)} = \mathbf{w}_j^{(k)} + \eta \sum_{i=1}^N \mathbf{b}(\mathbf{v}_i^{(k)}(\mathbf{w}))^T \frac{\partial \mathbf{v}_i^{(k)}(\mathbf{w})}{\partial \mathbf{w}_j}. \quad (2.21)$$

在大多数实际应用中, 用户知道一些正确的分配, 在这种情况下, 半监督方法变得更加合适。我们的算法可以很容易地适应这种半监督的情况, 通过自然地结合监督和非监督的学习步骤: 来自每个步骤的离散解  $\mathbf{b}$  具有固定的分配值, 对于这些分配, 地面真实信息是可用的, 而对于其余的未标记的分配, 我们使用, 在非监督的情况下, 由图匹配算法返回的解。我们提出的方法的另一个优点是能够以有原则的方式容易地将监督案例与非监督案例相结合。

## 2.8 学习超图匹配

最近的超图匹配方法[22–24]是基于张量  $\mathbf{H}$  的秩一近似(ROA)。 $\mathbf{H}$  的 ROA 是基于分配是概率独立的假设。让我们定义概率赋值向量  $\mathbf{p}$ , 其长度与候选赋值列表  $\mathbf{L}$  相同, 其中  $\mathbf{p}_{ia} \in [0, 1]$  是来自  $\mathbf{I}_m$  的特征  $\mathbf{I}$  与来自  $\mathbf{I}_t$  的特征  $\mathbf{a}$  匹配的概率。在统计独立性假设下, 张量  $\mathbf{H}$  可以通过以下使用外积的分解来近似[23, 24]:

$$\mathbf{H} \approx \prod_{i=1 \dots 3} \otimes \mathbf{p}. \quad (2.27)$$

---

#### Algorithm 2.3 Symmetric higher order power method (S-HOPM)

---

```

 $\mathbf{v} \leftarrow \mathbf{1}$ 
repeat
   $\mathbf{v}_{ia} \leftarrow \sum_{j,b,kc} H_{ia;jb,kc} \mathbf{v}_{jb} \mathbf{v}_{kc}$ 
   $\mathbf{v} \leftarrow \mathbf{v} / \|\mathbf{v}\|$ 
until convergence
return  $\mathbf{v}$ 

```

---

$\mathbf{p}$  的近似  $\mathbf{v}$  只有在三阶分数为  $H_{ia}$  时才起作用； $\mathbf{JB}; \mathbf{kc}$  正确地反映了潜在问题的结构和相似性。我们建议学习使  $\mathbf{v}$  和二元解向量  $\mathbf{b}$  之间的点积最大化的参数  $\mathbf{w}$  (等式 2.5)。  $\mathbf{b}$  的元素等于已知赋值和超图匹配算法为未知赋值产生的基本真值。这个想法扩展了先前章节和现在章节中的图形匹配学习方法。 [19]. 学习期间最大化的目标函数是训练图像对集合上每个  $\mathbf{v}$  和相应  $\mathbf{b}$  之间的点积之和:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \sum_i \mathbf{b}_i^T \mathbf{v}_i(\mathbf{w}). \quad (2.28)$$

要注意,  $\mathbf{v}(\mathbf{w})$  是参数  $\mathbf{w}$  的函数, 因为它取决于张量  $\mathbf{H}$  的元素, 这些元素都是  $\mathbf{w}$  的函数。我们通过梯度上升更新来优化目标:

$$w_j \leftarrow w_j + \eta \sum_{i=1}^N \mathbf{b}_i(\mathbf{w})^T \frac{\partial \mathbf{v}_i(\mathbf{w})}{\partial w_j}. \quad (2.29)$$

S-HOPM 的每次迭代计算更新  $\mathbf{v} \leftarrow \mathbf{v}/\|\mathbf{v}\|$ 。由此可见, 迭代  $k + 1$  时  $\mathbf{v}$  估计值的梯度可以表示为  $\text{gradient at the previous step}$  的函数。因此, 梯度可以与  $\mathbf{v}$  的计算一起进行计算。基于这些思想的递归算法 (算法 2.4) 可以基于这些计算思想导出。

---

**Algorithm 2.4** Compute the gradient  $d\mathbf{v}$  of  $\mathbf{v}$ 


---

```

 $\mathbf{v} \leftarrow \mathbf{1}$ 
 $d\mathbf{v} \leftarrow \mathbf{0}$ 
 $d\mathbf{H} \leftarrow \frac{\partial \mathbf{H}}{\partial \mathbf{w}_j}$ 
repeat
   $h_{ia} \leftarrow \sum_{jb;kc} dH_{ia;jb;kc} v_{jb} v_{kc} + 2H_{ia;jb;kc} v_{jb} dv_{kc}$ 
   $v_{ia} \leftarrow \sum_{jb;kc} H_{ia;jb;kc} v_{jb} v_{kc}$ 
   $\mathbf{h} \leftarrow \frac{\mathbf{h}}{\|\mathbf{v}\|}$ 
   $\mathbf{v} \leftarrow \frac{\mathbf{v}}{\|\mathbf{v}\|}$ 
   $d\mathbf{v} \leftarrow \mathbf{h} - (\mathbf{v}^T \mathbf{h}) \mathbf{v}$ 
until convergence
return  $d\mathbf{v}$ 

```

---

## 2.9 图匹配实验

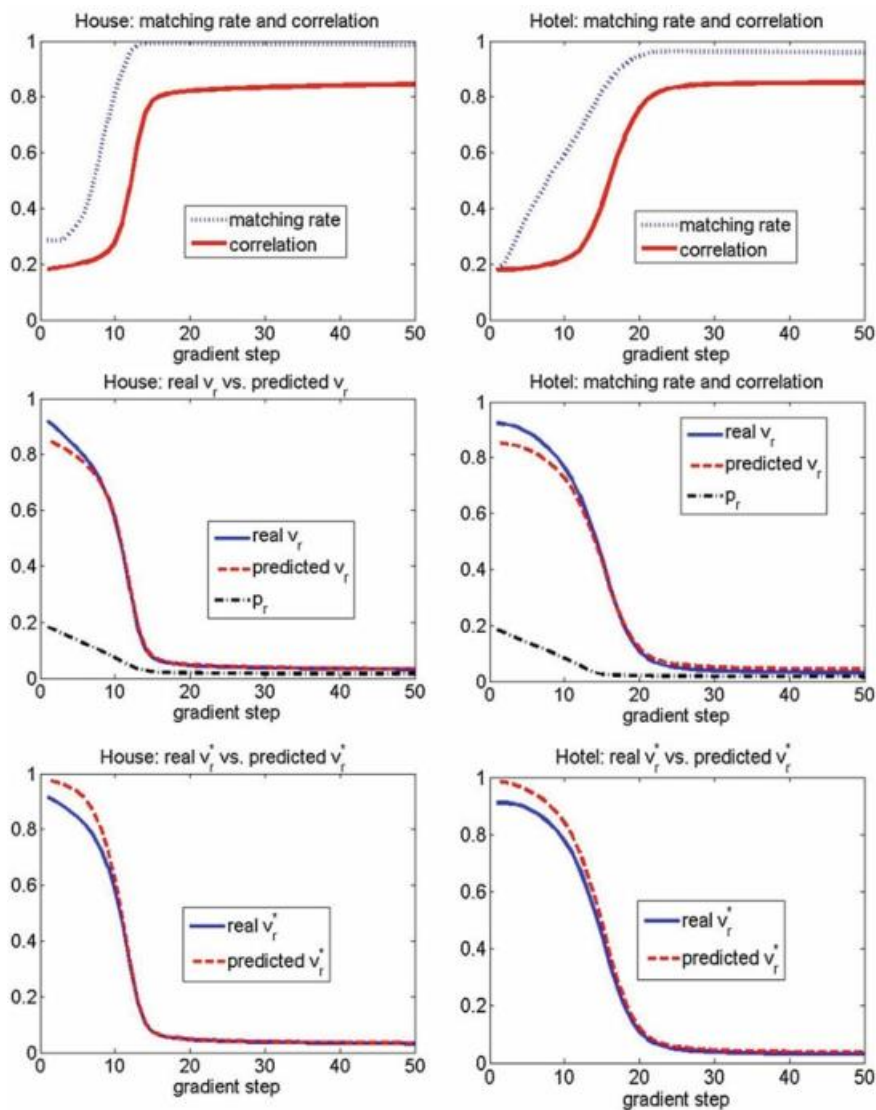
在图形匹配的情况下, 我们关注两个目标。第一个是验证第 2.7.1 节的理论结果, 特别是等式 2.12 和等式 2.13, 等式 2.12 建立了  $\mathbf{pr}$  和  $\mathbf{vr}$  之间的关系, 等式 2.13 将  $\mathbf{v}^*$  与  $\mathbf{vr}$  和误差率  $m$  联系起来。每个  $\mathbf{pr}$  是从训练序列上的每个单独的矩阵  $\mathbf{M}$  中凭经验估计的, 类似地, 每个  $\mathbf{vr}^*$  和  $\mathbf{VR}$  来自每个单独的特征向量。等式 2.12 很重要, 因为它表明, 与不正确分配之间的成对协议相比, 正确分配之间的成对协议越有可能 (由  $\mathbf{pr}$  选择), 匹配性能越好。

这个等式也验证了我们的匹配矩阵  $M$  的模型，它由两个平均值  $p_0$  和  $p_1$  分别定义。等式 2.13 很重要，因为它解释了为什么通过最大化相关  $v^T b(v)$  (并且隐含地最小化  $v^T \Sigma$ ) 我们实际上最小化了  $VR$  和匹配误差  $m$ 。等式 2.13 基本上显示了为什么无监督算法相对于基本事实确实将最大化性能。我们提到匹配率/性能是指正确匹配的特征的比率(在匹配特征的总数中)，而错误率是  $1 - \text{匹配率}$ 。

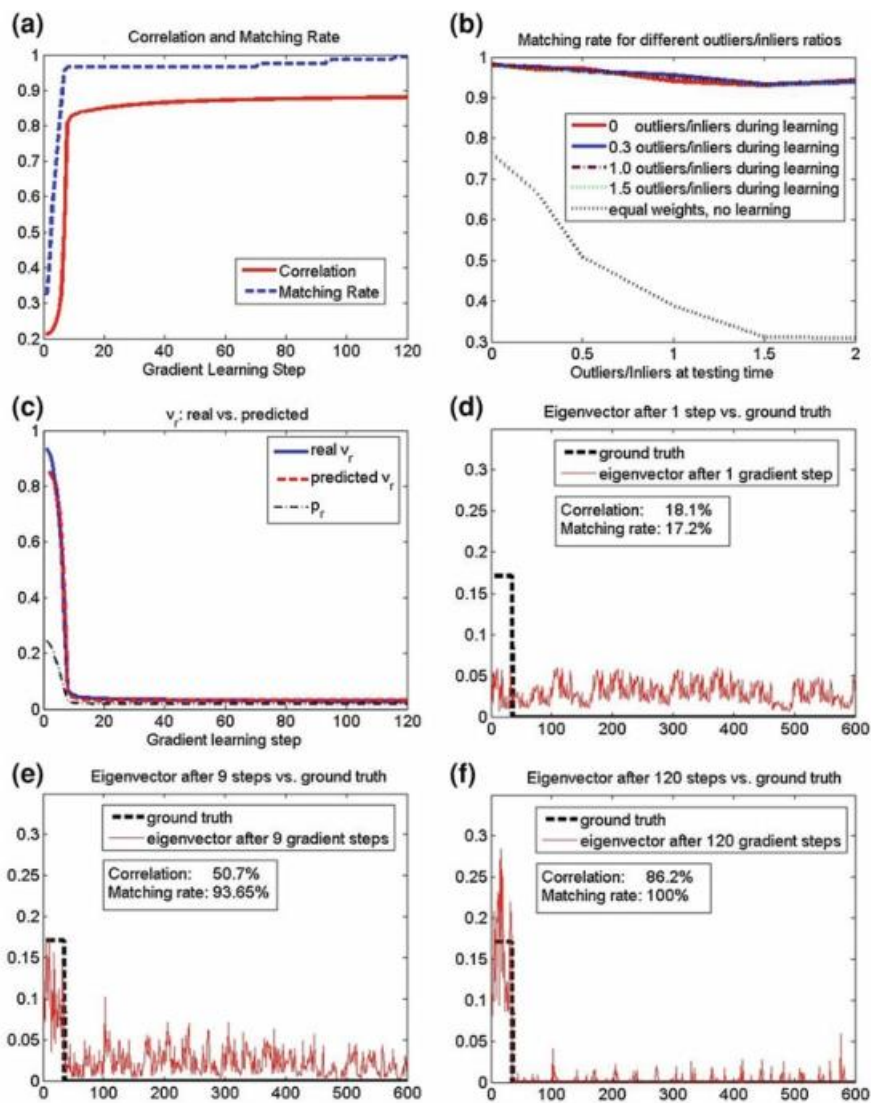
图 2.6、图 2.7、图 2.8 和图 2.9 分别显示了房屋、酒店、人脸、汽车和摩托车实验的结果，验证了我们的理论主张。这些实验的细节如下。

有几个相关的结果需要考虑。在所有四个不同的实验中， $v$  和地面真实值  $t$  之间的相关性在每个梯度步长都增加，即使地面真实值对于学习算法是未知的。匹配率与相关性同时以相似的速率提高，表明最大化这种相关性也会最大化最终性能。在图 2.7 中，我们显示了特征向量的一个典型例子，因为在训练过程中，例子越来越多。如果在第一次迭代之后，特征向量几乎是平坦的，但是在最后一次迭代中，所有正确的赋值都比任何错误的赋值具有更大的置信度。此外，在所有单个实验中，方程 2.12 和 2.13 的两个近似值随着每个梯度步长而变得越来越精确，从第一次迭代的不到 10% 的精度到最后的不到 0.5% 的误差。在我们所有的学习实验中，我们从一组不支持任何赋值的参数  $w$  开始 ( $w = 0$ ，这意味着在第一次迭代之前， $M$  中所有非零的分数都等于 1)。这些结果不仅激发了为  $M$  提出的模型(等式 2.12)，还激发了支持无监督学习方案的结果(等式 2.13)。

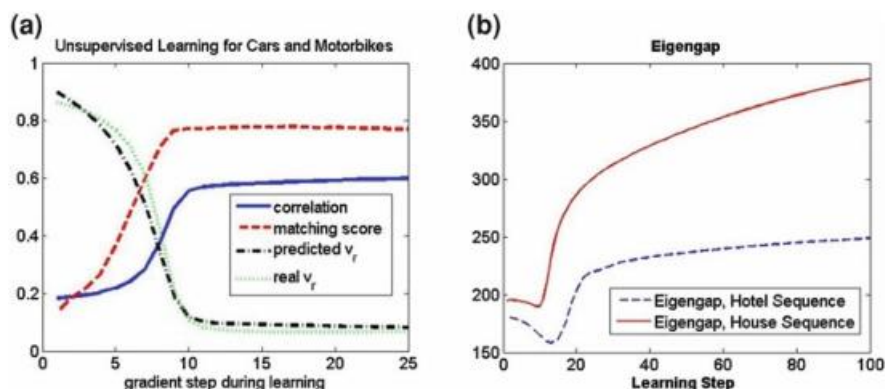
我们实验的第二个目标是评估学习前后新测试图像对的匹配性能。目标是表明，在测试时，学习后的匹配性能明显优于没有学习的情况。



**Fig. 2.6 无监督学习阶段。**第一行:在每个梯度步骤的训练期间,特征向量与地面事实的匹配率和相关性。剩下的图显示了方程 2.12 和 2.13 的左侧,即从每个图像对获得的特征向量经验估计的  $v_r$  和  $VR\Sigma$ , 判断它们的预测值是否一致(方程 2.12 和 2.13 的右侧)。结果是 70 个不同实验的平均值



**Fig. 2.7 人脸上的结果:**特征向量和背景真值之间的相关性, 以及训练期间的匹配率(左上), 测试时的匹配率, 学习和测试时不同的离群值/内联值比率(右上), 验证等式 2.12(左中), 不同学习步骤的示例特征向量。前三个图的结果是 30 个不同实验的平均值。



**Fig. 2.8 相关和匹配率。**帕斯卡 2007 挑战赛汽车和摩托车无监督学习的基本事实。将真实值和模型预测值进行比较。结果是 30 个不同实验的平均值。**b** 在无监督学习过程中，归一化本征间隙(本征间隙除以  $M$  中的均值)经过几次迭代后开始增大，表明前导本征向量越来越稳定。结果基于平均超过 70 次随机实验的房屋和酒店数据集

## 2.9.1 用未标记的对应关系学习

### 2.9.1.1 透视变换下的刚性对象匹配

我们首先对两个任务进行实验，这两个任务与 Caetano 等人的实验相同。[17]和我们以前的工作[18]。我们使用完全相同的图像序列(房屋:110 幅图像和酒店:100 幅图像)进行训练和测试，并使用由卡埃塔诺等人手动选择的相同特征。为了测试，我们使用所有的剩余图像。两两得分  $M_{ia; j b}$  与我们之前在 Leordeanu 和 Hebert [18]中使用的相同，使用形状上下文描述符[88]来表示外观，使用 *airwise distance* 和 *handogles* 来表示二阶关系。他们测量一幅图像的特征  $(I, j)$  在几何和外观方面与候选对应  $(a, b)$  的一致性。更明确地说，成对分数的形式是  $M_{ia; j b} = \exp(-w^T g_{ia; j b})$ ,  $g_{ia; j b} = [s_i - s_a, |s_j - s_b|, |d_{ij} - d_{ab}|, |\alpha_{ij} - \alpha_{ab}|]$ . 这里， $s_a$  表示特征  $a$  的形状上下文； $d_{ij}$  是特征之间的距离  $(I, j)$ ； $\alpha_{ij}$  是水平轴和矢量之间的角度。学习包括找到在训练序列上最大化匹配性能的参数向量。

正如卡埃塔诺等人。[17]和 Leordeanu 和 Hebert [18]，我们首先获得一个 Delaunay 三角剖分并允许非零成对分数  $M_{ia; j b}$  当且仅当  $(I, j)$  和  $(a, b)$  在它们对应的三角剖分中都是连通的。我们以前的方法[18]是有监督的，并且基于比本章中提出的无监督的基于梯度的方法更有可能找到真正的全局最优的全局优化方案。因此，值得注意的是，拟议的。

**表 2.1 匹配运行时间和暂停时间.** 在实验中, 我们只使用了来自“房子”序列的 5 个训练图像, 而对于来自卡埃塔诺等人的方法。[17], 我们使用 5 和 106 个训练图像报告了他们公布结果的上限。符号: “S” 和 “U” 分别表示 “有监督” 和 “无监督”

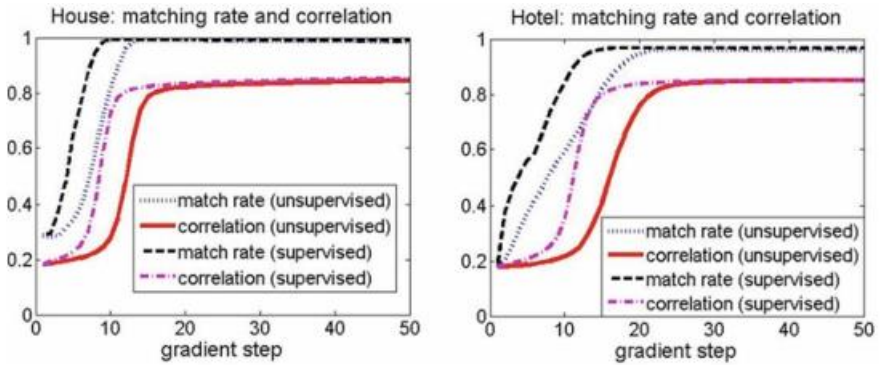
Dataset	Ours: S(5) (%)	Ours: U(5) (%)	Caetano et al. [17]: S(5) (%)	Caetano et al. [17]: S(106) (%)
House	99.8	99.8	<84	<96
Hotel	94.8	94.8	<87	<90

无监督学习方法匹配我们以前的结果, 同时显著优于 Caetano 等人的方法。[17](表 2.1)。

我们指出, 我们和卡埃塔诺等人在性能上有显著差异的主要原因。[17]与事实有关, 即[17]在成对分数 Mia 中较少强调二阶几何; jb, 仅使用 0–1 Delaunay 三角测量的信息, 不使用关于成对距离和角度的信息。相反, 我们强调二阶关系的重要性, 因为在我们的实验中, 即使我们完全忽略了形状上下文描述符和二阶几何信息, 我们的方法的性能也不会下降。当然, 同样重要的是要强调, 虽然卡塔诺等人的方法。[17]以有监督的方式学习参数, 我们是第一个以无监督的方式这样做的。

接下来, 我们研究了无监督学习方法在学习和测试阶段的性能与有监督学习方法的性能(当基本事实分配已知时)。我们使用这两个数据集进行了 70 个不同的实验, 随机选择 10 个训练图像(并使用训练集中的所有图像对), 并将其余图像的所有对留下来进行测试。正如预期的那样, 我们观察到无监督的方法平均比有监督的方法学习得慢一些, 但是参数学习者基本上是相同的.在图 2.9 中, 我们绘制了所有训练对和所有实验的每个梯度步长的平均相关性(特征向量和地面真实值之间)和匹配率与每个梯度步长的关系, 包括监督和非监督情况。有趣的是, 虽然无监督版本往往收敛较慢, 但经过几次迭代后, 它们的性能(以及参数)会收敛到相同的值。在测试期间, 这两种方法在匹配性能方面表现相同(在所有 70 个实验中正确匹配的特征的平均百分比)。与没有学习参数的相同匹配算法相比, 这两种算法的性能明显更好(表 2.2)。在没有学习的情况下, 默认参数(w 的元素)被选择为全部相等。





**Fig. 2.9 2.9 监督学习与非监督学习:**在所有训练图像对上的特征向量和地面真实之间的平均匹配率和相关性，分别使用来自豪斯和酒店序列的 10 个随机选择的训练图像的 70 多个不同实验。注意无监督学习如何收敛到与有监督学习相同的相关性和匹配率

**Table 2.2 在 70 个不同实验(10 个训练图像，其余用于测试)的房屋和酒店数据集上测试时平均匹配性能的比较。**我们比较了无监督学习和无监督学习的情况。第一栏:无监督学习；第二:不学习，等于默认权重  $w$

Datasets	Unsupervised learning	No learning
House + Hotel (%)	99.14	93.24

2.9.1.2 用异常值匹配可变形二维图形

第三个用于评估的数据集由从 Caltech-4 人脸数据集中选取的 30 对随机图像对组成。I 在这个数据集上的实验与之前的不同有两个原因:图像中不仅包含了人脸，还包含了背景杂乱。而且，这些脸属于不同的人，包括男人和女人，有着不同的面部表情，所以在脸之间有显著的非刚性变形，必须进行匹配。我们使用的特征是沿着轮廓提取的定向点采样，与我们之前的工作类似[50(图 2.10)]。每一点考虑的方向是该特定点处轮廓的法向量。人脸上必须匹配的点(inliers)是手动选择的，而离群点(背景中的特征)是随机选择的，同时要确保每个离群点与任何其他点都不会太近(15 像素)。对于每一对面孔，我们手动为 inliers 选择地面真实值(正确匹配)。成对分数只包含成对距离和角度的几何信息:

$$M_{ia; jb} = e^{-wT \mathbf{g}_{ia; jb}},$$

(2.30)



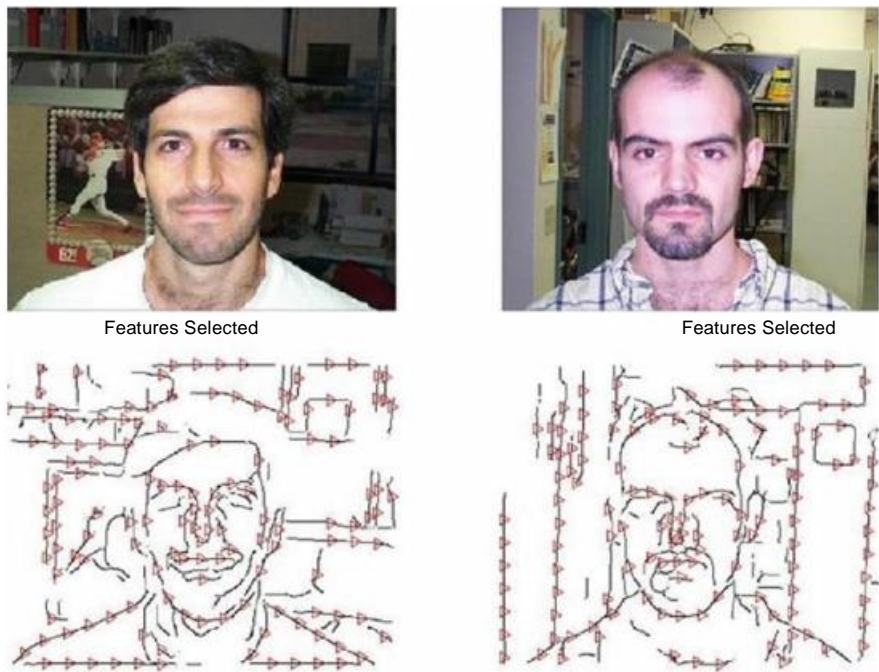


Fig. 2.10 顶行:我们实验中使用的加州理工学院 4 号数据集的一对人脸。底行:提取的轮廓和选择的点作为特征

这里  $w$  是有 7 个参数的向量,  $g_{ia;jb} = [|d_{ij} - d_{ab}|/d_{ij}, |\theta_i - \theta_a|, |\theta_j - \theta_b|, |\sigma_{ij} - \sigma_{ab}|, |\sigma_{ji} - \sigma_{ba}|, |\alpha_{ij} - \alpha_{ab}|, |\beta_{ij} - \beta_{ab}|]$ . 这里,  $d_{ij}$  特性之间的距离( $i, j$ )是正常的夹角特性和水平轴,  $\sigma_{ij}$  是正常的向量夹角  $\rightarrow ij$ ,  $\alpha_{ij}$  之间的角度是  $\rightarrow ij$  和水平轴, 和  $\beta_{ij}$   $i$  和  $j$  的法线之间的夹角。

我们进行了 30 个实验, 随机挑选 10 对进行训练, 剩下的 20 对进行测试。图 2.7 所示的结果是 30 个实验的平均值。左上角的图显示了在之前的实验中, 训练中的相关  $vTt$  和匹配性能是如何随着每个学习步骤而提高的。在训练和测试期间, 我们使用不同百分比的异常值来评估方法的稳健性(右上图)。这种学习方法对异常值是鲁棒的, 因为测试期间的匹配性能不依赖于在训练期间引入的异常值的百分比(异常值的百分比总是与主图像的百分比相同), 而只依赖于测试时出现的第三方地理因子。在没有学习的情况下(黑色虚线), 当选择的默认参数都相等时, 性能会差得多, 并且随着测试时异常值百分比的增加, 性能会下降得更快。这表明, 学习不仅提高了匹配率, 而且使其对异常值的存在更加鲁棒。

### 2.9.2 不同图形匹配算法

我们对帕斯卡 2007 年挑战赛的汽车和摩托车数据进行了最新的图表匹配算法。2.11)，与[19]中使用的相同。我们使用与前面描述的相同的特征(从轮廓片段中选择的定向点)。这些算法是谱匹配[16] (SM)、仿射约束谱匹配[3] (SMAC)、分级赋值[1] (遗传算法)、概率图匹配[24] (概率模型)和我们的整数投影不动点[6] (IPFP)算法。

对于每种算法，我们为每一个类执行 30 个不同的学习和测试实验，并对结果进行平均。对于每个实验，我们随机挑选 10 对图像进行学习(有异常值)，剩下的 20 对进行测试(有异常值和没有异常值)。在训练过程中，我们在每对图像中的一个图像上添加离群点，这样离群点与内联点的比率为 0.5。该对图像中的另一个图像不包含异常值。我们在训练期间引入这种中等数量的离群值，以便在真实世界的实验中测试无监督学习方法的鲁棒性，其中，特别是在无监督的情况下，在每对图像中的两个图像中实施相等数量的特征是耗时的。在测试过程中，我们有两种情况：在第一种情况下，我们在两幅图像中都没有异常值，而在第二种情况下，我们只允许在一幅图像中出现所有可能的异常值。引入的离群值的数量是显著的，离群值与内联值的比率对于汽车类从 1.4 到 8.2 不等(平均为 3.7)，对于摩托车类从 1.8 到 10.5 不等(平均为 5.3)。根据基本事实，作为一个例外的其他测试，我们指的是在另一个图像中存在对应关系的异常值，而一个异常值没有这样正确的对应关系。内联线是手动选取的，方式与前面实验中使用的相同，而外联线是在轮廓上随机选取的，因此没有外联线与任何其他选定的特征之间的距离超过 15 个像素。

在图 2.12 中，我们显示了每个算法在学习过程中的行为：平均匹配率和每个学习步骤中特征向量与基础事实的平均相关性。有几个重要的方面需要注意：对于所有算法，特征向量和地面真实之间的相关性随着每个梯度步长而增加，SMAC 收敛得比其他算法快得多。这也反映在 SMAC 的匹配率增长更快。所有算法都受益于学习，因为经过几次迭代后，所有匹配率都有显著提高。参数  $w$  的向量被初始化为零，最终学习的  $w$  对于所有算法都是相似的。遗传算法和 SMAC 在最初的 10 步中有快速的改进，随后是一个小的下降和剩余迭代的平稳。这可能表明，对于遗传算法和 SMAC 学习应该只执行少数迭代。对于其他三种算法，学习在训练过程中不断提高匹配率。

在表 2.3 中，当引入异常值时，显示了两个数据集的所有算法的测试结果，包括学习和不学习。不学习所有算法，在两个数据集上使用一个参数向量  $w = [0.2, 0.2, 0.2, 0.2, 0.2]$ 。在我们的实验中，我们在测试中引入的异常值越多，学习就变得越有益。这也是一个没有经验的问题(图 2.7)。表 2.3 显示了在存在大量异常值的情况下学习和不学习的结果。



Fig. 2.11 帕斯卡 2007 挑战赛图像对的匹配结果。请注意形状、视点和比例的显著差异。最佳彩色视图

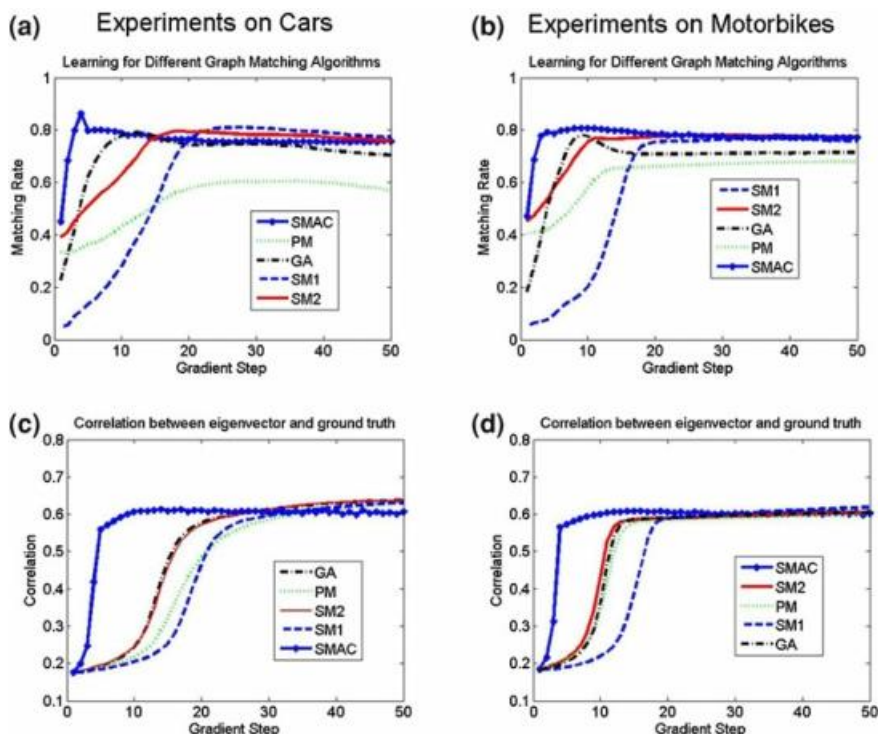


Fig. 2.12 第一行:在每个梯度步骤的每个算法的训练期间,相对于地面事实的平均匹配率。第二行:在每个梯度步骤的每个算法的学习期间,主特征向量和基本事实之间的平均相关性。SMAC 比其他算法收敛得更快。所有算法学习的最终参数都是相似的。

(如前所述,一幅图像中没有异常值,另一幅图像中有所有可能的异常值)。很明显,学习可以显著提高所有算法的性能。本节显示的结果强烈表明,在存在大量异常值的情况下,我们的无监督学习方案可以显著提高其他算法在困难数据上的性能(例如帕斯卡'07的汽车和摩托车)。

### 2.9.3 条件随机场实验

为了将我们的方法与以前在 CRFs 上的工作进行比较,我们严格遵循了 Kumar 关于图像去噪的实验,遵循了实现细节。

表 2.3 不同图形匹配算法在测试时的匹配率的比较，在(NL)之前和之后(WL)未监督的警告和摩托车公司的 07 年数据库中，有异常值:在右侧图像中允许所有异常值，在左侧图像中没有异常值。测试时使用的算法与学习时使用的算法相同。结果是 30 个不同实验的平均值。所有的算法都使用相同的参数，不管有没有异常值:w 的所有元素都是相等的

Dataset	IPFP (%)	SM (%)	SMAC (%)	GA (%)	PM (%)
Cars (NL)	<b>50.9</b>	26.3	39.1	31.9	20.9
Cars (WL)	<b>73.1</b>	61.6	64.8	46.6	33.6
Motorbikes (NL)	32.9	29.7	<b>39.2</b>	34.2	26.1
Motorbikes (WL)	<b>55.7</b>	54.8	52.4	46.8	42.2

和测试数据在 Kumar [89]中提供。任务是从损坏的二进制  $64 \times 64$  图像中获得去噪图像。我们使用了相同的四幅图像和相同的噪声模型。对于更简单的任务，噪声模型是高斯模型，均值  $\mu = 0$ ，标准偏差  $\sigma = 0.3$  添加到 0–1 二进制图像中。对于更困难的任务，我们像在库马尔[89]中一样，对每个类别使用两个高斯混合的不同混合物，它们具有相等的混合权重，产生双峰噪声。两个高斯分布的模型参数(均值，标准差)分别为前景类[(0.08, 0.03)，(0.46, 0.03)]和背景类[(0.55, 0.02)，(0.42, 0.10)].图 2.13 显示了原始图像及其噪声版本的例子。

与库马尔的方法[89]不同，库马尔的方法使用 50 幅随机生成的第一幅图像的噪声版本进行训练，我们只使用了 5 幅这样的图像。对于更简单的任务，我们还执行了完全无监督的学习，得到了几乎相同的结果(表 2.4)。对于更简单的噪声模型，我们的结果明显更好，而对于更困难的噪声模型，我们的结果与 Kumar [89]的结果相匹配。此外，请注意，我们的学习方法简化了性能改进(图. 2.14)，而不仅仅是 L2QP(我们从 Leordeanu 和 Hebert [37]中得到的谱图推断算法，最初就是为它设计的)。我们使用的成对势是

$$M_{ia; jb} = \sigma(\mathbf{w}^T [t_a; t_a I_i; t_b; t_b I_j; t_a t_b |I_i - I_j|]),$$

(2.31)

其中， $I_i$  是图像中像素  $I$  的值， $|I_i - I_j|$  是连接位置  $I$  和  $j$  之间图像像素值的绝对差值。在 Kumar [89]之后，我们使用了四连通格。我们还对八个连通的邻域进行了实验，它们在性能上没有显著差异。





Fig.2.13 第一行:原始二进制图像(左一个用于培训, 下三个用于测试).第二行:图像被单峰噪声破坏。第三行:图像被双峰噪声破坏

Table2.4 实验中与库马尔[89]的比较.InKumar[89], 第一幅图像的 50 个版本用于训练。我们只使用了第一幅图像的五个噪声版本进行训练。为了测试这两种方法, 使用其余三个图像的 50 个噪声版本。请注意, 无监督学习与有监督学习的性能相匹配。Kumar [89] 中使用的推理方法是图割, 学习方法是最大伪似然(PL)和最大惩罚伪似然(PPL)

Algorithm	L2QP (%)	IPFP (%)	Kumar [89]: PPL (%)	Kumar [89]: PL (%)
Unimodal (sup.)	0.75	<b>0.73</b>	2.3	3.82
Unimodal (unsup.)	0.85	<b>0.69</b>	NA	NA
Bimodal (sup.)	7.15	15.94	<b>6.21</b>	17.69

我们所有学习实验的参数值是

- 开始,  
 $\mathbf{w} = [0.5; -1; 0.5; -1; -0.5; 1];$
- 无监督的单峰噪声,  
 $\mathbf{w} = [1.27; -2.55; 1.27; -2.55; -2.50; 0.47];$
- 受监督的单峰噪声,  
 $\mathbf{w} = [1.27; -2.55; 1.26; -2.55; -2.63; 0.26];$
- 受监督的双峰噪声 ,  
 $\mathbf{w} = [1.98; -5.24; 1.98; -5.24; -2.99; 0.22].$

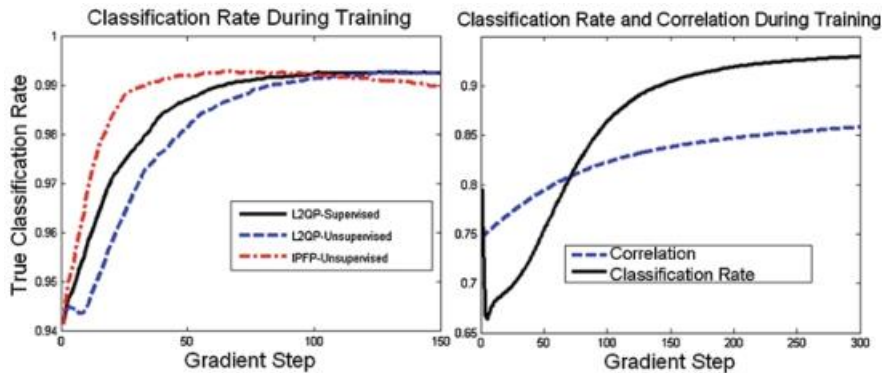


Fig.2.14 左图:使用单峰噪声时的监督和非监督学习。右图:使用双峰噪声时的监督学习。结果是每个学习梯度步骤的五个训练图像的平均值

我们的学习方法避免了大多数概率方法的计算瓶颈，如最大似然和伪似然，它们需要归一化函数  $z$  的估计。无监督学习在地图问题上不如在图形匹配上做得好的主要原因是矩阵  $M$  的不同结构。在图匹配的情况下，该矩阵包含主要由正确分配形成的单个强聚类，而在映射问题的情况下，该矩阵可以包含对应于完全不同标签的几个这样的聚类。偶然对齐的思想不适用于大多数地图问题，因此学习算法可以收敛到几个参数向量，将连续解二值化，在这种情况下需要监督学习。此外，即使在监督学习的情况下，在映射问题的情况下，训练对初始化也很敏感，这也是其他研究人员观察到的事实。

2.11 结论和今后的工作

在这一章中，我们提出了基于二次整数规划公式的图/超图匹配和映射推理问题的高效学习和优化方法.我们通过大量广泛的实验证明，这些算法有可能显著提高匹配差异的准确性。我们为图和超图匹配领域带来了有价值的见解和方法，为进一步的探索开辟了新的道路，特别是在高效匹配和无监督学习方面。我们处理了匹配的两个主要方面:(1)在连续域中找到松弛问题的最优解，以及(2)获得具有重要理论性质的高质量离散解，不同于基于贪婪算法或经典匈牙利方法获得最终硬分配的传统方法。

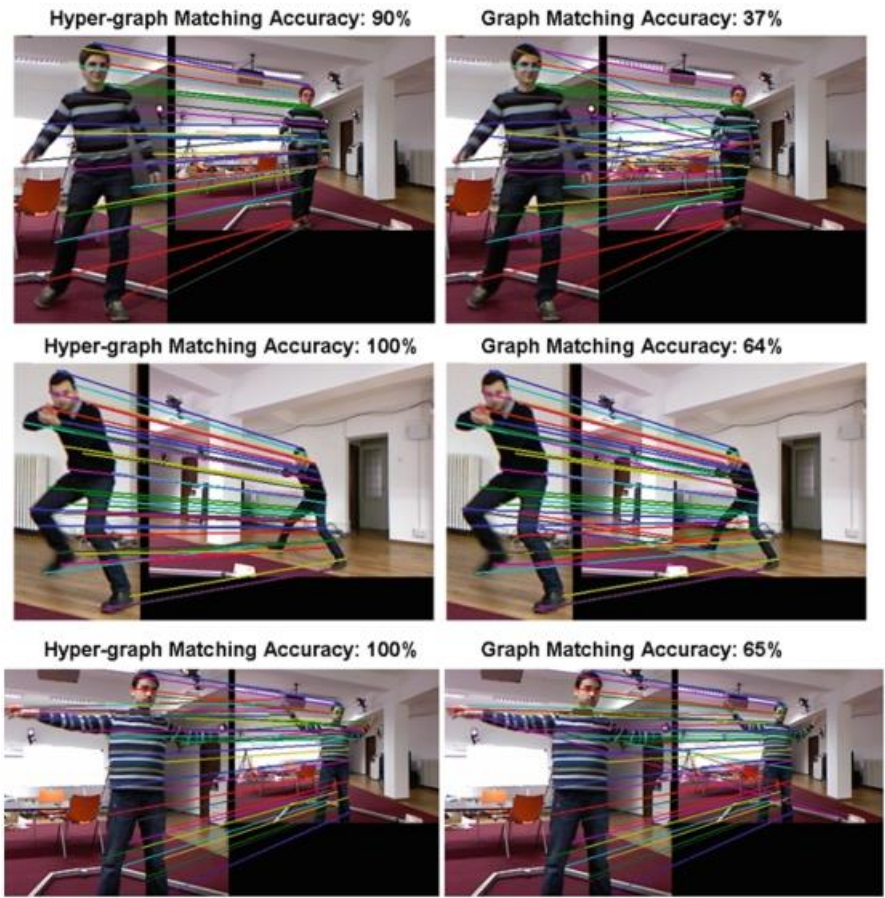


Fig.2.20 不同比例和姿势的人搭配例子。注意尺度的显著变化和身体部位相对较大的位移

在匹配二阶或高阶约束方面，有许多未来工作的途径。它们大大提高了匹配性能，因为它们对比例、旋转、透视或仿射变换的变化以及各种外观变化和非刚性变形更加鲁棒。例如，请参见图 2.23，在全局仿射变换的情况下，超图匹配在匹配精度方面如何容易地优于众所周知的、流行的使用 SIFT 特征的匹配方法。

下面列举了一些我们想进一步探讨的问题和方向。





Fig. 2.21 匹配不同尺度和姿势的人。请注意比例的显著变化和身体部位相对较大的位移

### 2.11.1 高效优化

我们的光谱匹配算法的引入促进了基于图形匹配的计算机视觉应用的发展，从医学成像、2D 和 3D 配准和重建、对象跟踪、建模、发现和学习。图形匹配的应用被扩展到更高层次的解释，如视觉识别和视觉理解。我们认为在未来的应用中，二阶或更高阶几何关系和范畴关系的使用将是非常重要的。按照这种思路，要解决的主要技术问题之一将是效率和计算成本。尽管光谱匹配很快，但作为一种优化方法，

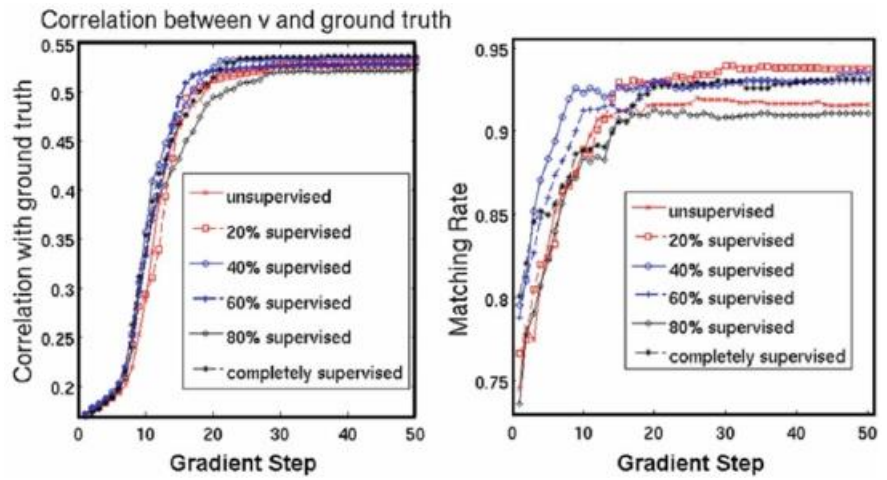


Fig.2.22 在培训期间改变监督的程度。请注意，平均而言，监管程度和收敛速度之间几乎没有相关性。对于每个监督程度，随机选择训练图像对

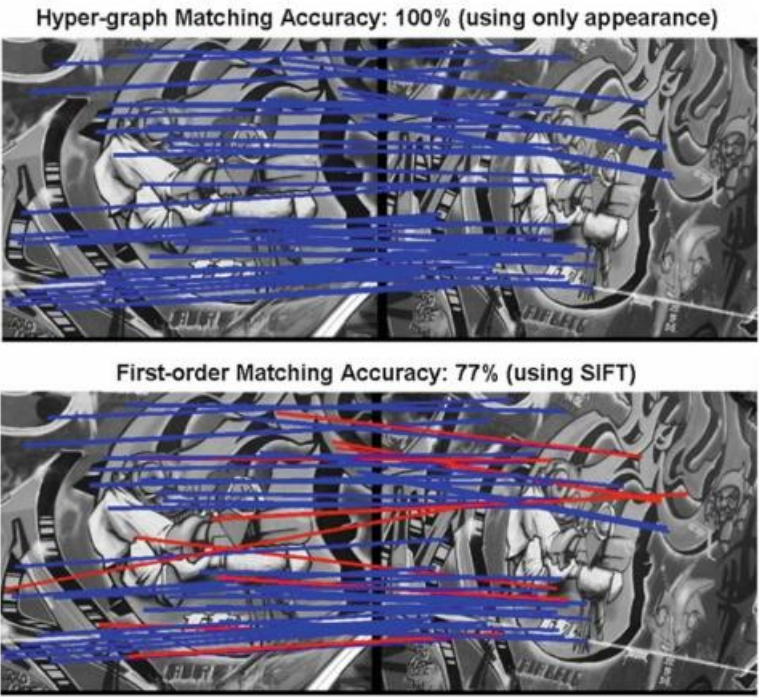


Fig. 2.23 仿射变换下的超图匹配与 SIFT 特征匹配

赋值矩阵或张量(对于高阶情况)的速度很慢。我们认为,需要在这方面取得进一步的实质性进展。我们未来的计划包括更深入地研究这个计算方面,研究如何在不损失准确性的情况下尽可能快地构造/建立问题。神经网络深度学习的成功影响了许多其他领域,其中许多深度方法目前正在被提出。我们并不惊讶地看到深度学习和卷积神经网络也被应用于图形匹配的增长趋势[21, 91–97]。

事实上,最近的光流估计方法,DeepFlow [98]以及我们的递归图神经网络模型和更一般的物体识别系统在第二章中介绍。8 也深深植根于使用图形进行匹配的模式。

随着当前深度卷积网络模型的能力开始变得更加清晰和易于理解,经典图形模型随着人口的不断增加而改变也就不足为奇了。标准的深度学习方法只有一个前馈通道,没有返回和纠正错误的可能性,以确保达到平衡。寻找不同处理阶段之间的平衡解通常涉及使用迭代消息传递过程,该过程自然会收敛到极限解。在更经典的概率图形模型和图形匹配方法中,这种迭代方法是学习和推理的核心。因此,可以预期,在不久的将来,我们将会看到越来越多的方法将前馈深层网络的优势结合起来——能够学习强大的功能,与经典的迭代图方法相结合,这有利于达成一致的解决方案。

### 2.11.2 高阶关系

正如我们对三阶匹配的实验所显示的,高阶项比二阶项更强大。然而,这种好处是以很大的代价获得的:很难找到具有良好匹配精度的快速近似优化算法,而设置问题、建立张量甚至需要更多的计算和内存。这是为什么高阶匹配没有像图匹配那样成功的主要原因之一,尽管它在理论上很有吸引力。[81],我们提出了基于局部仿射模型的优雅的封闭形式解的强有力的运动和遮挡的估计,它将更高阶的几何项减少到更容易计算和优化的二次能量。我们相信,在未来的几年里,我们将见证致力于高阶匹配、检测和推理的研究的惊人发展,重点是高效近似和高阶计算模型。

这种能量的智能学习或手动设计应该考虑外观、几何、分割和遮挡线索,以及更高级别的分类信息,以便在现实世界的图像中实现强大的性能。在下一阶段,在图形匹配中还有许多其他方向可以探索,包括建模、理论和优化特性,以及实际效率和应用。想法很多,应用它们的可能方法甚至更多。随着计算机视觉领域的飞速发展和从研究到工业、计算机视觉和机器人技术的快速转变,研究人员将有真正的机会为这个相对年轻和激动人心的领域开发许多有趣和有用的算法。

## 参考文献

1. Gold S, Rangarajan A (1996) A graduated assignment algorithm for graph matching. *PAMI* 18(4):377–388
2. Berg A, Berg T, Malik J (2005) Shape matching and object recognition using low distortion correspondences. In: *CVPR*
3. Cour T, Srinivasan P, Shi J (2006) Balanced graph matching. In: *NIPS*
4. Schellewald C, Schnorr C (2005) Probabilistic subgraph matching based on convex relaxation. In: *International conference on energy minimization methods in computer vision and pattern recognition*
5. Torr P (2003) Solving Markov random fields using semidefinite programming. In: *International conference on artificial intelligence and statistics*
6. Leordeanu M, Hebert M, Sukthankar R (2009) An integer projected fixed point method for graph matching and map inference. In: *NIPS*
7. Cho M, Lee J, Lee KM (2010) Reweighted random walks for graph matching. In: *ECCV*
8. Egozi A, Keller Y, Guterman H (2013) A probabilistic approach to spectral graph matching. *Pattern Anal Mach Intell* 35(1)
9. Zhou F, la Torre FD (2012) Factorized graph matching. In: *Computer vision and pattern recognition*
10. Cho M, Lee K (2012) Progressive graph matching: making a move of graphs via probabilistic voting. In: *Computer vision and pattern recognition*
11. Tolias G, Avrithis Y (2011) Speeded-up, relaxed spatial matching. In: *ICCV*
12. Escolano F, Hancock E, Lozano M (2011) Graph matching through entropic manifold alignment. In: *Computer vision and pattern recognition*
13. Torresani L, Kolmogorov V, Rother C (2013) A dual decomposition approach to feature correspondence. *Pattern Anal Mach Intell* 35(2)
14. Liu Z, Qiao H, Yang X, Hoi S (2014) Graph matching by simplified convex-concave relaxation procedure. *Int J Comput Vis* 1–18
15. Zhou F, la Torre FD (2013) Deformable graph matching. In: *Computer vision and pattern recognition*, pp 2922–2929
16. Leordeanu M, Hebert M (2005) A spectral technique for correspondence problems using pairwise constraints. In: *ICCV*
17. Caetano T, Cheng L, Le Q, Smola AJ (2007) Learning graph matching. In: *ICCV*
18. Leordeanu M, Hebert M (2008) Smoothing-based optimization. In: *CVPR*
19. Leordeanu M, Sukthankar R, Hebert M (2009) Unsupervised learning for graph matching. *IJCV* 96(1)
20. Cho M, Alahari K, Ponce J (2013) Learning graphs to match. In: *International conference on computer vision*, pp 25–32

21. Zanfir A, Sminchisescu C (2018) Deep learning of graph matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2684–2693
22. Duchenne O, Bach F, Kweon I, Ponce J (2009) A tensor-based algorithm for high-order graph matching. In: CVPR
23. Chertok M, Keller Y (2010) Efficient high order matching. *IEEE Trans Pattern Anal Mach Intell* 32:2205–2215
24. Zass R, Shashua A (2008) Probabilistic graph and hypergraph matching. In: CVPR
25. Çeliktutan O, Wolf C, Sankur B, Lombardi E (2014) Fast exact hyper-graph matching with dynamic programming for spatio-temporal data. *J Math Imaging Vis* 1–21
26. Wang J, Chen S, Fuh C (2014) Attributed hypergraph matching on a Riemannian manifold. *Mach Vis Appl* 25(4)
27. Yan J, Zhang C, Zha H, Liu W, Yang X, Chu SM (2015) Discrete hyper-graph matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1520–1528
28. Nie WZ, Liu AA, Gao Z, Su YT (2015) Clique-graph matching by preserving global and local structure. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4503–4510
29. Liu AA, Nie WZ, Gao Y, Su YT (2016) Multi-modal clique-graph matching for view-based 3d model retrieval. *IEEE Trans Image Process* 25(5):2103–2116
30. Adamczewski K, Suh Y, Mu Lee K (2015) Discrete Tabu search for graph matching. In: Proceedings of the IEEE international conference on computer vision, pp 109–117
31. Jiang B, Tang J, Ding C, Luo B (2017) Binary constraint preserving graph matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4402–4409
32. Shi X, Ling H, Hu W, Xing J, Zhang Y (2016) Tensor power iteration for multi-graph matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5062–5070
33. Yan J, Cho M, Zha H, Yang X, Chu SM (2015) Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Trans Pattern Anal Mach Intell* 38(6):1228–1242
34. Leordeanu MD (2010) Spectral graph matching, learning, and inference for computer vision. PhD Thesis, Carnegie Mellon University
35. Leordeanu M, Zanfir A, Sminchisescu C (2011) Semi-supervised learning and optimization for hypergraph matching. In: IEEE international conference on computer vision
36. Leordeanu M, Zanfir A, Sminchisescu C (2011) Semi-supervised learning and optimization for hypergraph matching. In: ICCV
37. Leordeanu M, Hebert M (2006) Efficient map approximation for dense energy functions. In: International conference on machine learning
38. Cour T, Shi J (2007) Solving Markov random fields with spectral relaxation. In: International conference on artificial intelligence and statistics
39. Ravikumar P, Lafferty J (2006) Quadratic programming relaxations for metric labeling and Markov random field map estimation. In: International conference on machine learning
40. Yan P, Khan SM, Shah M (2008) Learning 4D action feature models for arbitrary view action recognition. In: CVPR
41. Gaur U, Zhu Y, Song B, Roy-Chowdhury A (2011) A string of feature graphs model for recognition of complex activities in natural videos. In: International conference on computer vision, pp 2595–2602
42. Liu J, Luo J, Shah M (2009) Recognizing realistic actions from videos “in the wild”. In: CVPR
43. Nayak N, Sethi R, Song B, Roy-Chowdhury A (2011) Motion pattern analysis for modeling and recognition of complex human activities. In: Guide to video analysis of humans: looking at people
44. Zhu Y, Nayak N, Gaur U, Song B, Roy-Chowdhury A (2013) Modeling multi-object interactions using string of feature graphs. *Comput Vis Image Underst* 117(10)
45. Kim G, Faloutsos C, Hebert M (2008) Unsupervised modeling of object categories using link analysis techniques. In: CVPR