指导教师：杨涛　　提交时间：2016/3/17

# CVPR2015 Paper Translation

No: 1
姓名：裴晓煊
学号：2013302580
班号：10011305

# 无人监督的视觉与相似图形

Fatemeh Shokrollahi Yancheshmeh, Ke Chen, and Joni-Kristian Kam ¨ar ¨ainen ¨
Department of Signal Processing, Tampere University of Technology, Finland
{fatemeh.shokrollahiyancheshmeh, ke.chen, joni.kamarainen}@tut.fi

## 摘要

语义上有意义的视觉模式的对齐，例如对象类，对于许多应用是一个重要的预处理步骤，如目标检测和图像分类。考虑到花在监督对齐方法的注释上的昂贵的人力，无人监督的定位技术是更有利的特别是对大规模问题。微调是可以有效且高效地实现图像冻结的方法，但它们需要适度的初始化，否则在实践中，很大程度上是无效的。对齐的视觉例如大角度的变化仍然是一个开放的问题。基于功能的方法可以在某种程度上解决这个问题，但需要手动选择好种子图像而且忽略了一个事实，那就是一个语义类的不同实例可以在视觉上非常不同（如哈利—戴维森和小轮摩托车都是摩托车）。在这项工作中，我们通过快速近似和非线性优化解决了广义分配问题，并在此基础上通过定义视觉相似性克服了上述缺点。从双向图像相似之处我们构建一幅图用于实现步进式排列，“变形”，以及一个图像到另一个的过渡。我们会自动找到一个合适的种子来测量标识“相似性中心”图。该方法优于最先进的基于常规的基准数据集和类无监督的方式方法。

## 1. 引言

视觉图像的对齐的关键是找到它们之间的相应的控制点。对于一系列高层次的计算机视觉应用程序来说这是一个重要的预处理步骤，例如对

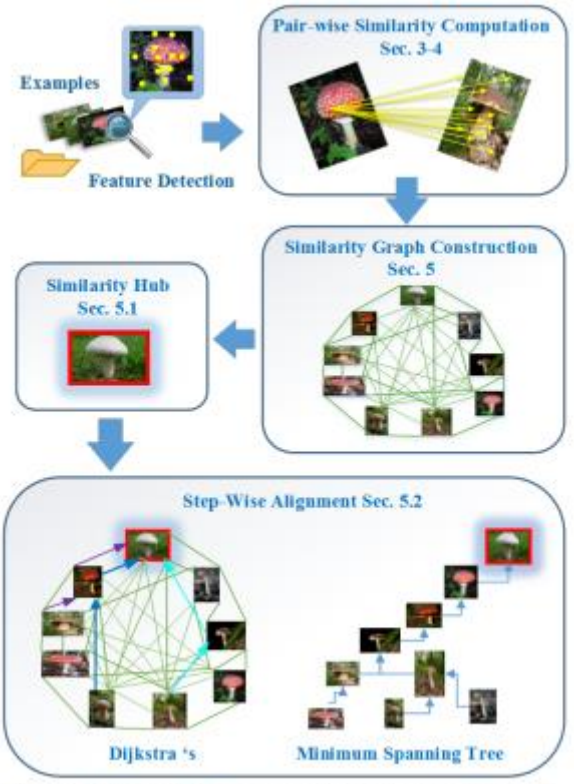


Figure 1. The workflow of our visual alignment approach.

象的检测和分类[1, 11, 15, 23, 37]。在那些应用程序中，图像中的实例类的大型姿势变化会造成很大困难。对齐仍就是具有挑战性的话题特别是对于以下大规模视觉识别问题：

i）在监督对象的对齐中，避免人力用于注释控制点或对象的坐标[46]；ii）缺乏图像冻结算法所需的适度的初始的排列[16]；iii）基于特征的对齐中的人工种子选择[26]。

最近的“大可视数据”数据库，如ImageNet [38]和 COCO [27]，使仅以类标签这种最小的注释培训有效的类探测器成为可能，从而避免过度学习，尽管模型参数数量巨大[24]。目前还不清楚这是否是由于类的量子化，但某些不良属性表明这个问题并没有完全解决[40]。基于部分的方法虽优于大型数据集和深度神经网络，但是限于目前的技术需要附加注释[10, 11, 25]。边界框有自己的缺点，最好的基于部分方法在一定程度上被黑客克服，如边界框集群[11]，确定离散子类或姿势[45]。黑客与有限的姿势工作,但失败在与类似的空间维度和遭受的对象子类之间不平衡的数据[1, 23]，很明显，更强的注释，如明确的对象姿势或地标改善检测。缺点是更广泛的人工注释是必要的。

另一种解决人工注释是无监督的视觉对齐对象类图像。准确定位可以有效地实现与最近的形象冷凝的方法[43 29 日,16 日,9 日,17）[29, 16, 43, 9,17]。然而,这些需要在实践中适度好初始化在很大程度上是无效的。另一种方法是最近基于对齐[26],然而,基于功能特性的方法的主要缺点是需要手动选择和搜索好的种子图像,目前仍依赖于人类专用的努力。此外,这两种方法都可能会失败在视觉子类下相同的语义标签,如摩托车、自行车运动和哈雷摩托车。

在这项工作中,我们采用基于功能特性的方法,但是克服上述缺点设计成对一个视觉相似性问题解决任务通过快速近似和非线性优化。从我们构建一幅图,双向的相似之处用于步进式排列,“变形”,一个图像到另一个图(见图 1)旅行。我们的方法也由小说中心自动找到一个合适的种子图中标识“相似性中心”。该方法优于非监督的方式从流行的最先进的方法和类基准数据集。源代码和数据对我们的实验将向公众公开。

## 2. 相关工作

我们的兴趣是在语义层面上的视觉对齐，因此我们省略对齐工作相关的不同的观点相同的场景(缝合) [4]和具体的方法类,比如面临[7, 34, 19]。特殊情况也包括跨域匹配[39],和非刚性的登记[5]和临时对齐[21]。

**图表表示**——一直被应用在各种作品的视觉对象匹配[20, 44, 36, 30]利用图结构来表示的一组图像。然而,通常使用一个图表只代表之间的空间配置的功能两个图片[20],或省略连接空间星座(Bag-of-Words)[44]或特定目的是匹配对象类[36,30]。作者的最好的知识,我们的工作是第一个模型成对相似性广义分配问题,表示“相似” 在接触相似图。

**图象冷凝**——语义水平对齐动量的开创性工作后 Learned-Miller[29]。其扩展[17,16,43,9]提供有效和高效精密对准后适度好的初始对准。冷凝的方法栈执行逐步转换优化堆栈相似图片。替换像素(如与当地特性。,筛选[28])提供了更好的鲁棒性成像失真[16、17]。一种已经提出的方法[26]是不需要初始对准是基于空间的验证地方特色[35]。这个问题在这个基于功能的冷凝是手动选择的好种子可能需要测试所有图片和缓慢的基于 RANSAC 空间验证。我们的方

法使用功能,但定义了广义相似任务框架。种子选择由我们解决中心测量和相似图。

## 3. 基于功能的相似性

在下一节中,我们使用术语"成本"尽管标题中使用的术语是"相似"。这种差异是故意的因为许多相关建立匹配的成本和工作，例如，欧几里得距离提供了一个直观的衡量或者他们的坐标匹配特性。然而，在建筑我们切换到相似的优化框架 4.2 节，我们提出我们的快速近似算法基于特征图像相似性。

我们基于特征相似(成本)是积极的部分原因表示成功地采用视觉类检测[12,25,11]。质量的特点对齐两个图像 Ia 和 Ib 测量在两个方面：类似的功能点对应的功能如何分，多少的空间排列的功能点改变。匹配的功能可分为特征匹配和特征几何失真成本

$$C(I_a, I_b) = \lambda_1 C_{match}(I_a, I_b) + \lambda_2 C_{dist.}(I_a, I_b) \ , \quad (1)$$

λ1 和 λ2 是两个术语之间的权衡参数。等价物(1)已经使用的对象从图像搜索匹配[3]，草图[2]和非刚性的匹配的图像序列工作[41]和[2]解决问题假设小几何扭曲和[3]需要手动分割存储类的原型。

我们的基于特征图像表示由 N 特征描述符 Fi = 1…N(如。,筛选[28])及其空间坐标 ξ= 1…N。匹配两个图像的成本(1)可以写成

$$C(I_a, I_b) = \lambda_1 C_{match}(F^{(a)}, F^{(b)}) + \lambda_2 C_{dist.}(\boldsymbol{x}^{(a)}, \boldsymbol{x}^{(b)}) \ . \quad (2)$$

形式分离特性匹配和几何失真,但是是误导,因为两者是相关的。成本函数的隐式地假定匹配的变量已知：$\boldsymbol{A}_{N_a \times N_b}$和 T 变换。赋值矩阵元素 aij 定义为$F_i^{(a)}$的特征值 Ia 对应于$F_j^{(b)}$的特征值 Ib。T 变换，如一个 3×3 线性单应性矩阵，从 Ib 的空间变换到 Ia 的空间。提供了证据的视觉外观特征匹配和 T 的几何畸变,因此更多成本的准确定义

$$\begin{aligned} &C(I_a, I_b) := C(I_a, I_b; \boldsymbol{T}, \boldsymbol{A}) = \\ &C\left(\left\{F^{(a)}, \boldsymbol{x}^{(a)}\right\}, \left\{F^{(b)}, \boldsymbol{x}^{(b)}\right\}; \boldsymbol{T}, \boldsymbol{A}\right) = \\ &\lambda_1 C_{match}\left(F^{(a)}, \boldsymbol{A}F^{(b)}\right) + \lambda_2 C_{dist.}\left(X^{(a)}, \boldsymbol{A}\boldsymbol{T}(X^{(b)})\right) \end{aligned} \quad (3)$$

特性匹配项成本取决于特性描述符和赋值和几何失真成本项取决于作业和变换 T(·)。特性匹配项成本取决于特性描述符和赋值，而几何失真成本项取决于作业和变换 T(·)。赋值可以通过矩阵乘法和转换的一个实际的例子是一个 T 3×3 单应性矩阵的变换 T(·)包括非齐次之间的映射和齐次坐标。

## 4. 相似度算法

### 4.1. 问题公式化

相似的成本(3)可以用来找到成对相似图片 Ia 和 Ib 的价值对于一个给定的几何变换 T a 实际作业然而,问题是

$$C(I_a, I_b; \boldsymbol{T}, \boldsymbol{A}) = \min_{\boldsymbol{T}, \boldsymbol{A}} C(I_a, I_b) \ .$$

通过定义参数变换 T 是一个问题，我们可以把问题最小化

$$\begin{aligned} \text{minimize} \quad & \sum_i \sum_j c_{ij} a_{ij} \\ \text{subject to} \quad & \sum_j a_{ij} \le 1 \ \ i = 1, \ldots, N_a \\ & \sum_i a_{ij} \le 1 \ \ j = 1, \ldots, N_b \\ & a_{ij} \in \{0, 1\} \end{aligned} \ , \quad (4)$$

在作业成本 cij = C(i,j)可以计算给定变换 T 和描述符 F(a)和 F(b),最直接的解决方案是采用 Na ×Nb 描述符和位置的距离

$$\boldsymbol{D}^F(i,j) = ||F_i^a - F_j^b||, \ \ \boldsymbol{D}^X(i,j) = ||\boldsymbol{x}_i^a - \boldsymbol{T}(\boldsymbol{x}_j^b)|| \quad (5)$$

然后将它们相加：

$$C = \lambda_1 \boldsymbol{D}^F + \lambda_2 \boldsymbol{D}^X \quad .$$

(4)形式的平凡解的全局最优$a_{ij} = 0 \ \forall i, j.$。为了避免平凡解的一个需求执行所有功能(Ia)映射到一些特性这是通过改变不平等平等：

$$\Rightarrow \sum_j a_{ij} = 1 \ \ i = 1, \ldots, N_a \quad . \tag{6}$$

尽管允许没有反馈功能和离群值，但人们还需引入 Na 和 F 代表离群值，即 Fa 的相应特征没有分配给任何的功能在 Fb，但假设离群值的固定成本 ε：

$$\begin{aligned} &F_j^b = F^\epsilon \text{ and } c_{ij} = \epsilon \\ &\text{for } i = 1, \ldots, N_a, j = N_b + 1, \ldots, N_b + N_a \end{aligned} \quad . \tag{7}$$

使用上面的扩展(4)减少作业量，用 O(N 3)解决问题，如匈牙利方法存在[33]。

作为一个可选择的解决方案，我们可以避免假作业和成本估计在(6)和(7)通过改变相似最小化成本 C 为最大化的相似性 S

$$\begin{aligned} &\text{maximize} \sum_i \sum_j s_{ij} a_{ij} \\ &\text{subject to} \sum_j a_{ij} \leq 1 \ \ i = 1, \ldots, N_a \\ &\qquad \sum_i a_{ij} \leq 1 \ \ j = 1, \ldots, N_b \\ &\qquad a_{ij} \in \{0, 1\} \end{aligned} \quad . \tag{8}$$

(8)的形式避免了哑变量和产量功能分配,提供了最大的相似性值(T)的转换。最大化问题是,称为广义分配问题这是 np 困难甚至 APX-hard 近似[6]。接下来我们介绍我们的快速近似最大化问题的计算复杂度 O(N)。

**4.2 近似**

我们可以映射的距离(成本)(5)(0,∞)相似度值在[0,1]的指数函数

$$\boldsymbol{S}(i,j) = e^{\boldsymbol{C}(i,j)} = e^{\lambda_1 \boldsymbol{D}^F(i,j)} e^{\lambda_2 \boldsymbol{D}^X(i,j)} = \boldsymbol{S}^F(i,j) \boldsymbol{S}^X(i,j) \quad .$$

**Algorithm 1** Generalized assignment approx. solution

1: Compute the feature distance matrix $\boldsymbol{D}^F_{N_a \times N_b}$ (*e.g.*, SIFT[28]).
2: On each row of $\boldsymbol{D}^F$ set the $K$ smallest to 1 and 0 otherwise.
3: $\boldsymbol{S}^X = \boldsymbol{0}$.
4: **for** $i = 1 : N_a$ (features of $I_a$) **do**
5: Compute the distance from $\boldsymbol{x}_i^{(a)}$ to $\boldsymbol{T}(\boldsymbol{x}_j^{(b)})$ for $j = 1, \ldots, K$ non-zero entries of $\boldsymbol{D}^F$ and if $\boldsymbol{D}^X(i,j) \leq \tau_X$ then set $\boldsymbol{S}^X(i,j) = 1$ and break.
6: **end for**
7: **return** the number of non-zero terms in $\boldsymbol{S}^X$

以指数形式 λ1λ2 定义精确匹配的相似性衰变。λ2 可以在空间中定义域与直观的解释，但定义 λ1 特征空间是很困难的,例如,结构筛选特征空间尚不清楚。

从计算的观点来看，在 SF 和 SX 或两者都接近于零的情况计算相似度值是浪费的。为了使相似矩阵稀疏化，我们用阶跃函数 H(·)(即单位阶跃函数,不连续函数的值为负参数,反之为正参数) 取代指数函数

$$\boldsymbol{S}(i,j) = \mathcal{H}(\boldsymbol{D}^F(i,j) - \tau_F)\mathcal{H}(\boldsymbol{D}^X(i,j) - \tau_X) \tag{9}$$

τF 和 τX 亥维赛阈值定义从 1 到 0 点相似之处。(9)的形式提供了大量加速的两个原因:

1．第一项并不依赖于 T 变换，但是是恒定的,可以提前和计算；

2．只有 SF 是非零的才需要计算 SX。为避免测量复杂特征空间我们取代功能距离 SF 与他们的等级次序的距离,例如输入 SF=1，τF = K 是 Ib 最好的特性匹配。这个设置相似矩阵 $S(i, j)$是 $KNa$ 的稀疏二进制零输入。为进一步加速计算，我们发现可以级联计算一个固定的空间阈值 τX 和第一点停止阈值。基于这些设置，最小数量的计算需要的是 Na 和最大 KN 那集近似分配解决方案的最终计算复杂度 O(N)。在实验中发现,K= 5,τX = 0.02 最好,此时 τX 分辨率独立除以与图像对角线的距离。完整的解决是用几行伪码算法 1 所示。

通过设置的类型变换 T- 2 d 相似,我们有四个自由度:翻译(x,y), φ 旋转和缩放(年代)。这只是一个四维搜索空间,我们能够有效地利用著名的 Nelder-Mead 非线性优化技术[31]。Nelder-Mead 的组合优化的基础上的快速近似广义分配问题提供了快速计算两两图像相似性的年代 $\boldsymbol{S}(Ia, Ib)$。

## 5．相似度图

使用前面定义的方法来计算两两图像的相似之处我们可以构建一个完整的 N×N 图像相似度矩阵

$$\boldsymbol{G}(i,j) = \boldsymbol{S}(I_i, I_j)$$

这是一个加权邻接矩阵与图 G 一个完整的连接。图表描述了视觉相似性从加州理工学院- 101 摩托车类的例子所示如图 2 所示。相似性计算使用近似算法基于不对称自等级次序最大化的年代 S(Ii,Ij)功能基数是有界的。在这个阶段我们不利用不

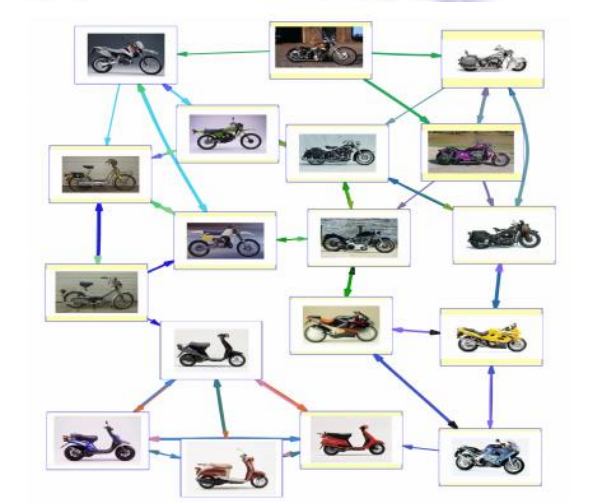

图 2。摩托车相似图由成对相似性年代(Ii,Ij)链接绘制最强(10%)。

对称属性但执行加权邻接矩阵对称的:

$$\boldsymbol{G}(i,j) = \max(\boldsymbol{G}(i,j), \boldsymbol{G}(j,i)).$$

大多数图形对称 G 算法,如最小生成树(MST)可用,但再次我们发现基于排序统计数据的表示有效的在我们的实验中:

$$\boldsymbol{G}(i,j) = \frac{N}{\text{rank}\,(\boldsymbol{G}(i,j), \text{sort\_ascend}(\boldsymbol{G}(:,j)))} .$$

值 1 表示强烈的距离(图)和 N 的值低相似性(距离长图)。与上述相似值变换加权邻接矩阵 G 代表一个接触无向图。

### 5.1. 相似的“枢纽”

由图表图形结构中心的定义，图像 G 包含节点,通常出现在两个随机图像之间的最短路径 *Ii* 和 *Ij*[13、14]。我们通过自动识别这些“校准中心”可以选择适合其他图像准确地对齐。这样的中心对应于手动选择“种子”[26]。

灵感来自于随机漫步亲密中心[32]定义一个中心测量基于第一和第二每个照片节点的顺序相似统计:

$$\mu_i = \frac{1}{N}\sum_j \boldsymbol{G}(i,j),\ \ \sigma_i = \frac{1}{N-1}\sqrt{\sum_j (\boldsymbol{G}(i,j) - \mu_i)^2} .$$

确定节点与其他节点异常相似，单节点统计比较在所有节点的平均统计数据

$$\mu = \frac{1}{N}\sum_i \mu_i,\ \ \sigma = \frac{1}{N-1}\sqrt{\sum_j (\mu_i - \mu)^2} .$$

我们选择一组中央枢纽 H 使用统计测试片面的正态分布:(1×方差对应 16%的最佳值):

$$H = \{I_i\} \text{ for which } \mu_i \geqslant \mu + \sigma .$$

选择一个图像 I0 从其余的 H 一致通过我们切换到二阶统计数据最小的节点相似度方差:

$$I' = \underset{I_i \in H}{\text{argmin}}\ \sigma_i .$$

### 5.2. 步进式对准中央枢纽

所有的图片都可以对齐到一个空间的

最中心的形象 I0 中心确定的计算提出了 5.1 秒。。可能的对齐策略利用图 G 结构如下:

•直接对齐[26]

•最小生成树(MST)路径(例如整洁算法[8])或

•图最短路径(例如 Djikstra 的算法[8])。

在我们的实验中第一个选项直接对齐提供业绩不佳,特别是在情况下输入图像包含的例子从视觉上不同的子类。在这些情况下,其他两个策略,利用图形旅行更有效,让视觉上不同的图像之间的“变形”,从“摩托车”到“HarleyDavidson”。结论在实验中得到了证明。

## 6. 实验

在我们的实验中我们使用相同的数据集和性能从最近的对齐和固定措施作品[26，17]的作者提供代码运行与上述基准的方法。此外,我们选择和具有挑战性的带注释的地标从 ImageNet 数据类。原则上,评估是基于手工注释地标哪个最好映射到相同的位置和对齐误差为零。然而,对于超过 3 地标不是准确的映射了保障,在这种情况下我们“理想的效果”表示最好的对齐的带注释的地标。在所有的实验中我们的描述符检测到使用密集的筛选 VLFeat 工具箱[42]。

### 6.1. 与目前最先进的比较

在我们的第一个实验中,运行使用所提供的数据和地标基准[26]。基准尤其适合基于对齐方法[26]和黄等不适合冷凝的方法艾尔。[17](随机加州理工学院- 101 年被引入的同一作者[22])。在图 3 的结果理想的对齐(手动地标),两种 state-ofthe-art 方法。x 轴是平均均方误差校准后的地标和 y 轴表示图像的数量(max.50)特定的对准精度。通常,在图像对角归一化条件的距离测量用于人脸检测的文献中，≤0.05 是优秀的，≤0.10 是好的，≤0.15 是令人满意的。值得注意的是,对我们所有实例,无监督方法优于基于功能特性的方法,最优的是手动选择。冷凝图像方法在人为制造偏差大幅平移、旋转和尺度变化时完全失败。

### 6.2. 计算负担

在这个实验中我们取代了 RANSAC 空间匹配算法 Lankinen et al。[26]与我们的快速分配算法和 Nelder-Mead 优化(4.2 秒)和比较的准确性和计算时间。结果相同的四类与前面的实验收集表 1。请注意,在我们的表有固定的操作点(轴)0.10 代表 良好的对准精度和报告的比例图像的精度。我们的方法是持续 2 - 3 倍,产生相同或更好的精度。结果验证我们的理论框架相似优化更有效和高效的启发式 RANSAC 匹配[26]。

### 6.3. 逐步调整

评估两个分段策略,Dijkstra 算法最短路径算法和拘谨的最小生成树(MST)算法(5.2 秒),我们比较了加州理工学院- 101 年,ImageNet 类。有趣的是,研究结果通常,而互补;图像偏差与 MSTDijkstra 算法正确一致,反之亦然。此外,发现与小几何方差 Dijkstra 算法更好的(在图 4),但与重要几何变化 MST 产生更好的对齐(海星)。一般来说,迪杰斯特拉是首选由于更可靠的结果。

除了逐步策略，直接可以使用对齐。然而,完全失败了在很多子类,非常敏感中心选择平均甚至在良好条件不如 MST 和 Dijkstra 算法逐步调整。那在图 5 演示了两个 ImageNet 类。注意，猫鼬包含 3 d 带来变

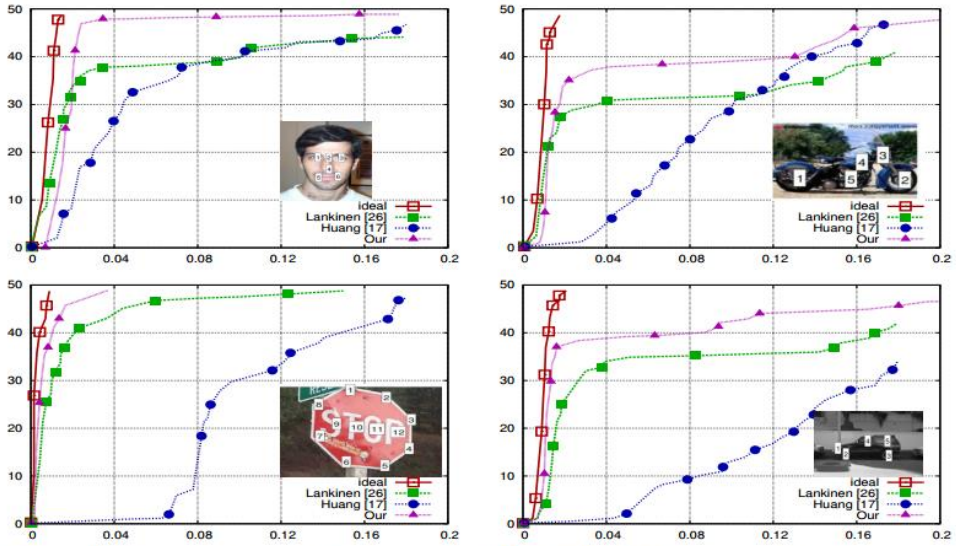


Figure 3. Our method vs. state-of-the-art for the benchmark in [26] (r-Caltech-101 Faces, motorbikes, stop sign and cars).

Table 1. The proportions of correctly aligned images (accuracy threshold 0.10) and the computation times for the feature-based alignment (FB) in [26] with the original RANSAC matching and with our fast assignment solver in Sec. 4.2.

| |  |  |  |  |
|---|---|---|---|---|
| FB [26] acc. | 86% | 76% | 98% | 74% |
| comp. time ($s$) | 170 | 61 | 163 | 96 |
| FB with Alg. 1 | 86% | 76% | 100% | 80% |
| comp. time ($s$) | 83 | 27 | 60 | 45 |

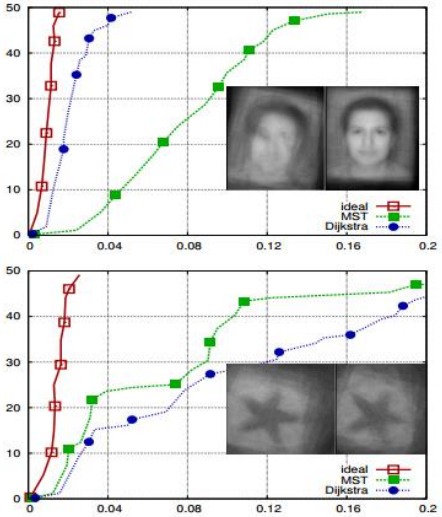


Figure 4. Comparison of the MST and Dijkstra algorithms for the stepwise alignment of Caltech faces and starfish images (left: average image of MST, right: Dijkstra).

化，最理想的结果显然更糟比前面的例子来自加州理工学院的数据集。

### 6.4. 面部验证

除了校准实验我们运行的面部验证实验从[17]标记的面孔在野外数据库[18]伦敦时装周开幕。因为大量的脸图像使数据集非常“密集”几何变换，我们基准更加困难通过随机抽样 150 名身份子集，子集进行训练和测试，和平均结果。面对边界框的自动 idenfitied 中心图像转换为其他面临图片使用最好的枢纽功能和在那之后我们运行匹配算法从[17]使用他们的代码(细节可以找到原来的文章)。结果未对齐，凝固的和我们的表 2 中,我们逐步对齐图像 Dijkstra 算法步进式对齐方法是最好的。

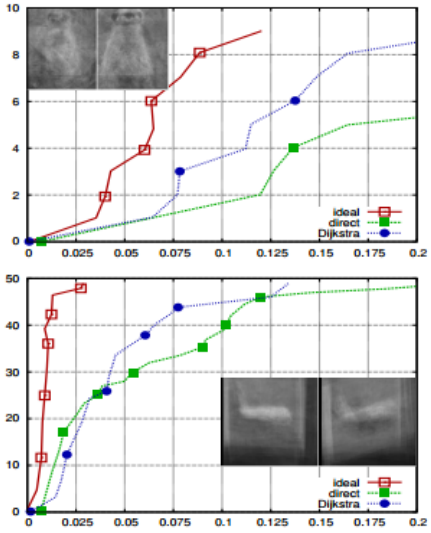


Figure 5. Direct vs. stepwise (Dijkstra) alignment of ImageNet classes airplanes and meerkat (left: direct, right: stepwise).

Table 2. The Labeled Faces in the Wild face verification benchmark [17] with aligned images.

| Alignment | Avg. Accuracy |
|---|---|
| Original images | 57.1% |
| Congealing [17] | 64.2% |
| Direct (Ours) | 53.6% |
| Dijkstra's (Ours) | 71.4% |

## 7. 结论

在这项工作中,我们研究的问题无监督图像校准和调整的图像对象类的例子。前面的冷凝方法和基于特征匹配遭受人工种子选择、缺乏良好的初始对准和视觉遥远子类。我们定义一个成对图像相似性度量,结合本地部分相似和几何失真（3 秒）和实际的相似度值发现通过搜索最大的相似性函数。搜索是广义的组合分配问题和非线性优化(4.1 秒)我们提出了一个有效的和高效的近似 4.2 秒。解决我们建造了一个接触相似子类和种子的选择问题图的种子被认定为“相似中心” (5.1 秒)所有图像可以使用对齐跳过“演变”图节点(5.2 秒)在实验中,我们的方法优于最先进的 semisupervised(人工选种)基于功能的方法和最先进的无监督图像冻结。

## 鸣谢

该项目是由芬兰科学院主持，项目批准号 267581,D2I SHOK，由 Digile Oy 和诺基亚技术(坦佩雷,芬兰)和博士学校博士学位授予由坦佩雷大学提供技术。作者感谢芬兰 CSC 科学中心提供慷慨的计算资源。

## 参考文献：

[1] H. Azizpour and I. Laptev. Object detection using stronglysupervised deformable part models. In *ECCV*. Springer, 2012. 1, 2

[2] S. Bagon, O. Brostovski, M. Galun, and M. Irani. Detecting and sketching the common. In *CVPR*, 2010. 2

[3] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *CVPR*, 2005. 2

[4] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *Int J Comput Vis*, 74(1), 2007. 2

[5] F. Brunet, V. Gay-Bellile, A. Bartoli, N. Navab, and R. Malgouyres. Feature-driven direct non-rigid image registration. *Int J Comput Vis*, 93, 2011. 2

[6] R. Cohen, L. Katzir, and D. Raz. An efficient approximation for the generalized assignment problem. *Information Processing Letters*, 100(4), 2006. 3

[7] T. F. Cootes, C. J. Twining, V. S. Petrovi, K. O. Babalola, and C. J. Taylor. Computing accurate correspondences across groups of images. *PAMI*, 32(11), 2010. 2

[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009. 5

[9] M. Cox, S. Sridharan, S. Lucey, and J. Cohn. Least squares congealing for large number of images. In *CVPR*, 2009. 2

[10] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008. 1

[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained partbased models. *PAMI*, 32(9), 2010. 1, 2

[12] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *Int J Comput Vis*, 61(1), 2005. 2

[13] L. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 1977. 5

[14] L. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1, 1979. 5

[15] E. Gavves, B. Fernando, C. G. M. Snoek, A. W. M. Smeulders, and T. Tuytelaars. Fine-grained categorization by alignments. In *ICCV*, 2013. 1

[16] G. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *ICCV*, 2007. 1, 2

[17] G. Huang, M. Mattar, H. Lee, and E. Learned-Miller. Learning to align from scratch. In *NIPS*, 2012. 2, 5, 6, 7

[18] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts,

Amherst, 2007. 7

[19] I. Kemelmacher-Shlizerman and S. Seitz. Collection flow. In *CVPR*, 2012. 2

[20] G. Kim, C. Faloutsos, and M. Hebert. Unsupervised modeling of object categories using link analysis techniques. In *CVPR*, 2008. 2

[21] G. Kim and E. Xing. Jointly aligning and segmenting multiple web photo streams for the inference of collective photo storylines. In *CVPR*, 2013. 2

[22] T. Kinnunen, J.-K. Kamarainen, L. Lensu, J. Lankinen, and H. Kalvi ¨ ainen. Making visual object categorization more ¨ challenging: Randomized Caltech-101 data set. In *ICPR*, 2010. 5

[23] I. Kokkinos and A. Yuille. Unsupervised learning of object deformation models. In *ICCV*, 2007. 1, 2

[24] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012. 1

[25] M. Kumar, A. Zisserman, and P. Torr. Efficient discriminative learning of parts-based models. In *ICCV*, 2009. 1, 2

[26] J. Lankinen and J.-K. Kamarainen. Local feature based unsupervised alignment of object class images. In *BMVC*, 2011. 1, 2, 5, 6

[27] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 1

[28] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision. The proceedings of the seventh IEEE international conference on*, volume 2, 1999. 2, 4

[29] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities of transforms. In *CVPR*, 2000. 2

[30] H. Myeong, J. Y. Chang, and K. M. Lee. Learning object relationships via graph-based context model. In *CVPR*, 2013. 2

[31] J. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7, 1965. 4

[32] J. Noh and H. Rieger. Random walks on complex networks. *Phys. Rev. Lett.*, 92(118701), 2004. 5

[33] C. Papadimitriou and K. Steiglitz. *Combinatorial Optimization*. Dover Publications, 2nd edition, 1998. 3

[34] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. In *CVPR*, 2010. 2

[35] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007. 2

[36] J. Philbin, J. Sivic, and A. Zisserman. Geometric latent dirichlet allocation on a matching graph for large-scale image datasets. *Int J Comput Vis*, 95(2), 2011. 2

[37] E. Riabchenko, J.-K. Kam ¨ ar ¨ ainen, and K. Chen. Learning ¨ generative models of object parts from a few positive examples. In *ICPR*, 2014. 1

[38] O. Russakovsky, J. Deng, Z. Huang, A. C. Berg, and L. FeiFei. Detecting avocados to zucchinis: what have we done, and where are we going? In *ICCV*, 2013. 1

[39] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros. Data-driven visual similarity for cross-domain image matching. In *SIGGRAPH Asia*, 2011. 2

[40] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. In *ICLR*, 2014. 1

[41] M. Torki and A. Elgammal. One-shot multi-set non-rigid feature-spatial matching. In *CVPR*, 2010. 2

[42] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008. 5

[43] A. Vedaldi and S. Soatto. A complexity-distortion approach to joint pattern alignment. In *NIPS*, 2006. 2

[44] S. Xia and E. R. Hancock. Incrementally discovering object classes using similarity propagation and graph clustering. In *ACCV*, 2009. 2

[45] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In *WACV*, 2014. 2

[46] X. Xiong and F. De la Torre Frade. Supervised descent method and its applications to face alignment. In *CVPR*, 2013. 1