

用户手册

Faceso FH1

面部表情及头部运动采集模块



心拓英启科技（北京）有限责任公司版权所有(2012-2021)

Braincraft Technology Co. Ltd

目 录

1. Faceso FH1 硬件介绍.....	2
1.1 硬件配置:	2
1.2 硬件安装调试.....	2
2. Faceso FH1 软件操作.....	2
3. Faceso FH1 数据说明.....	5
3.1 面部表情的构成.....	5
3.2 其他特征数据.....	8
4. Faceso FH1 面部表情算法简介.....	8
4.1 算法目标.....	8
4.2 面部模型.....	10
4.3 分类器.....	13
4.4 优化.....	14
4.5 初始化.....	14
4.6 计算性能.....	16
5.附件文献: 基于正则化的地标均值偏移方法拟合变形模型.....	17

1. Faceso FH1 硬件介绍

1.1 硬件配置：

名称	数量
摄像头	1
三脚架	2
补光灯	2
插线板	1

1.2 硬件安装调试

- 1、将网络摄像头固定于显示器后方支架上并连接至电脑，调整角度使其能完整的拍到被试面部。
- 2、将补光灯固定在三脚架上，然后将其调节至适当高度并放置被试座椅两侧。

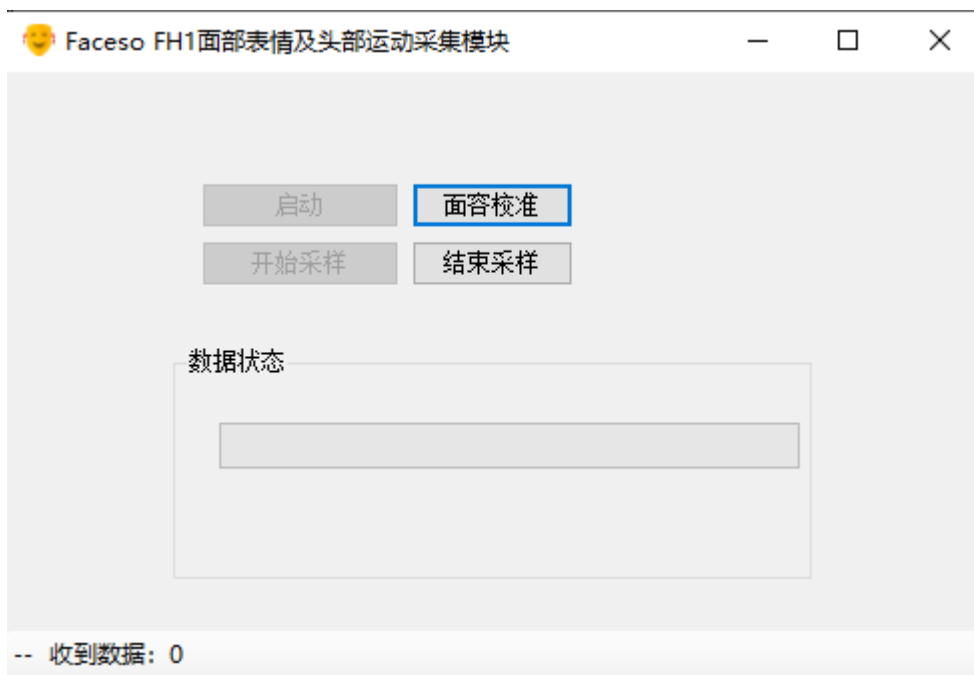
2. Faceso FH1 软件操作

1. 在 Eyeso Studio 眼动实验设计及数据分析系统中启动面部表情程序，只需在 Eyeso Studio 工具栏中点击“面部表情”图标即可。



2. 在Faceso FH1 面部表情及头部运动采集模块中点击开始采样,打开摄像头,

在面部表情识别网页中点击开始。



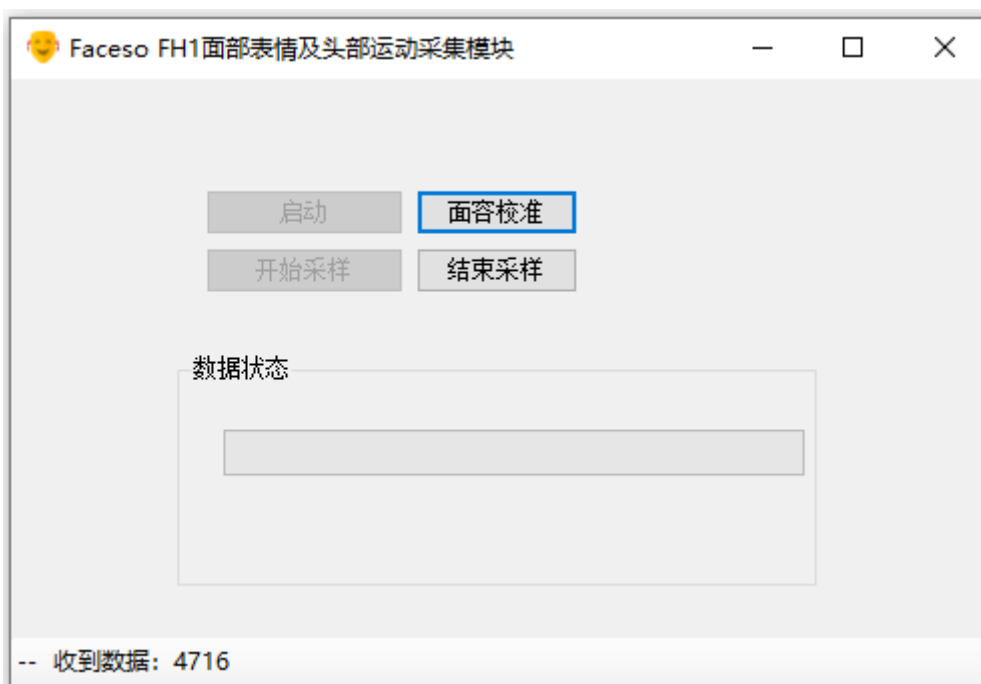
图：点击启动，开始采样



图：连接摄像头

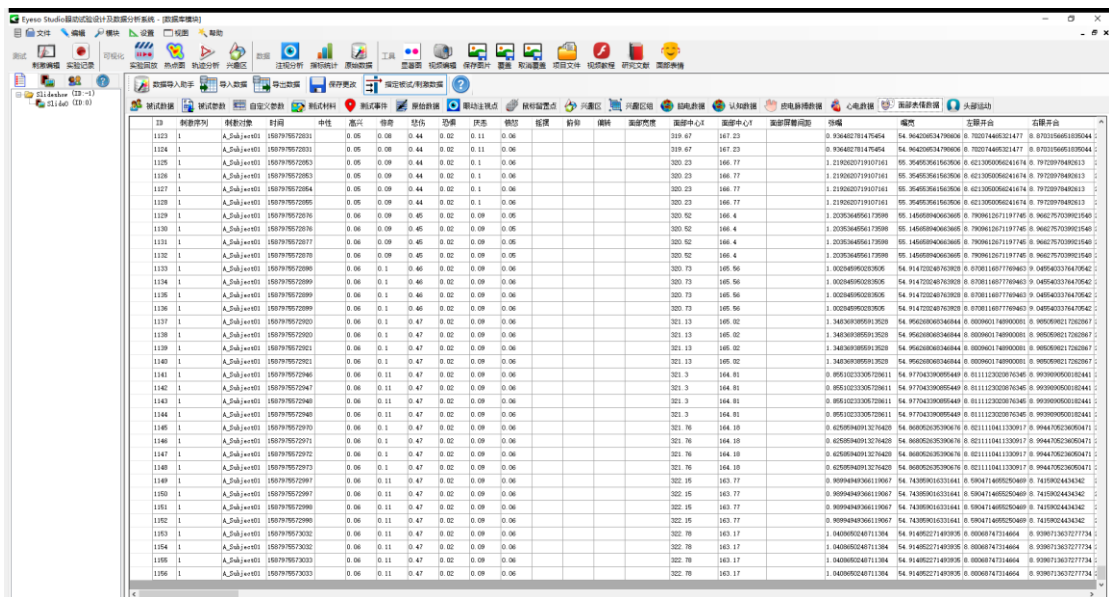


图：点击开始



图：左下角收到数据处，显示数据不为 0 即采样成功

- 面部校准，将补光灯打开调节到合适光源及摄像头光圈使软件能稳定识别面部为止。
- 在 Eyeso Studio 眼动实验设计及数据分析系统中按照记录眼动实验流程进行实验记录。
- 实验结束点击结束采样并关闭面部表情采集模块。
- 在原始数据模块可见面部表情数据。



3. Faceso FH1 数据说明

面部表情是指通过眼部肌肉、颜面肌肉和口部肌肉的变化来表现各种情绪状态。它是一种十分重要的非语言交往手段，通常可以通过面部表情来体现人物内心的情绪和情感。

3.1 面部表情的构成

眼：眼睛通常是情感的第一个自发表达者，透过眼睛可以看出一个人是欢乐

还是忧伤，是发恼还是悠闲，是厌恶还是喜欢，

眉：眉间的肌肉皱纹能够表达人的情感变化。柳眉倒竖便是愤怒，横眉冷对表示敌意，挤眉弄眼表示戏谑，低眉顺眼表示顺从，扬眉吐气表示畅快，眉头舒展表示宽慰，喜上眉梢表示愉悦。

嘴：嘴部表情主要体现在口形变化上。伤心时嘴角下撇，欢快时嘴角提升，委屈时撅起嘴巴，惊讶时张口结舌，忿恨时咬牙切齿，忍耐痛苦时咬住下唇。

面部：面部肌肉松弛表明心情愉快、轻松、舒畅，肌肉紧张表明痛苦、严峻、严肃。

1971年，Ekman 和 Friesen 研究了人类的 6 种基本表情（即高兴、悲伤、惊讶、恐惧、愤怒、厌恶），并系统地建立了人脸表情图象库，细致的描述了每一种表情所对应的面部变化，包括眉毛、眼睛、眼睑、嘴唇等等是如何变化的。

表 1：表情脸的运动特征具体体现：

表情	额头、眉毛	眼睛	脸的下半部
高兴	眉毛参考：稍微下弯	1下眼睑下边可能有皱纹，可能鼓起，但并不紧张 2鱼尾纹从外眼角向外扩张	1唇角向后拉并抬高 2嘴可能被张大，牙齿可能漏出 3一道皱纹从鼻子一直延伸到嘴角外部 4脸颊被抬起
惊奇	1眉毛抬起，变高变弯	1眼睛睁大，上眼皮抬高，	下颌下落，嘴张开，

	<p>2眉毛向下的皮肤被拉伸</p> <p>3皱纹可能横跨额头</p>	<p>下眼皮下落</p> <p>2眼白可能在瞳孔的上边和/或下边漏出来</p>	<p>唇和齿分开，但嘴部不紧张，也不拉伸</p>
悲伤	<p>眉毛内角皱在一起，抬高，带动眉毛下的皮肤</p>	<p>眼内角的上眼皮抬高</p>	<p>1嘴角下拉</p> <p>2嘴角可能颤抖</p>
恐惧	<p>1眉毛抬起来并皱在一起</p> <p>2额头的皱纹只集中在中部，而不横跨整个额头</p>	<p>上眼睑抬起，下眼皮拉紧</p>	<p>嘴张，嘴唇或轻微紧张，向后拉；或拉长，同时向后拉</p>
厌恶	<p>眉毛压低，并压低上眼睑</p>	<p>在下眼皮下部出现横纹，脸颊推动其向上，当并不紧张</p>	<p>1上唇抬起</p> <p>2下唇与上唇紧闭，推动上唇向上，嘴角下拉，纯轻微凸起</p> <p>3鼻子皱起</p> <p>4脸颊抬起</p>
愤怒	<p>1眉毛皱在一起，压低</p> <p>2在眉宇间出现竖直皱纹</p>	<p>1下眼皮拉紧，抬起或不抬起</p> <p>2上眼皮拉紧，眉毛压低</p> <p>3眼睛瞪大，可能鼓起</p>	<p>1唇有两种基本的位置：紧闭，唇角拉直或向下，张开，仿佛要喊</p> <p>2鼻孔可能张大</p>

六种基本表情的值都在 0-1 之间，代表确信度。比如厌恶 1 代表程序 100% 确认，厌恶 0.47 代表程序 47% 确认。某个时刻测出了不同表情的表情值，厌恶的表情值最大，代表这个时刻的表情是厌恶的可能性最大。

3.2 其他特征数据

张嘴、嘴宽、左眼开合、右眼开合、左眉抬升、右眉抬升分别记录了他们的运动状态。

P0-70 则记录某个时刻 71 个侦测像素点的坐标。

面部中心：相对于屏幕，面部中心点的坐标值。

4. Faceso FH1 面部表情算法简介

4.1 算法目标

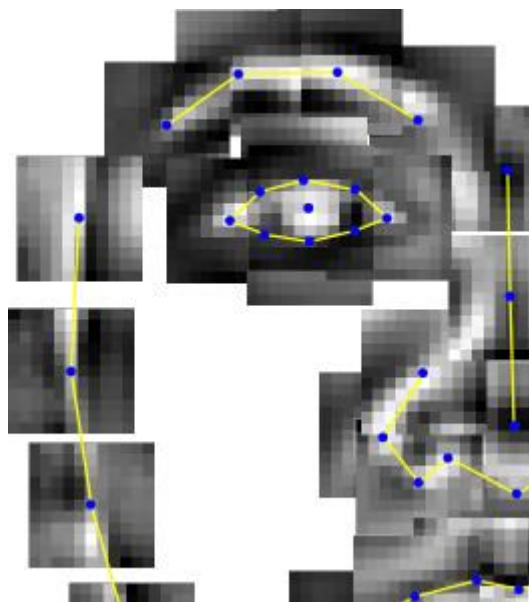
Faceso FH1 面部表情基于 Jason Saragih 和 Simon Lucey 在本文中描述的算法，更精确地说，是“通过子空间约束的均值偏移实现的人脸对齐”。

我们的目标是通过一个近似的初始化将一个面部模型应用到图像或视频中的一个面部。在我们的例子中，面部模型由 70 个点组成，如下所示。



面部模型

该算法采用 70 个小分类器，即模型中每一个点的一个分类器。给定一个初始的近似位置，分类器围绕每个点搜索一个小区域(因此命名为“local”)以获得更好的拟合，然后将模型增量地移动到给出最佳拟合的方向，逐渐收敛到最优拟合。

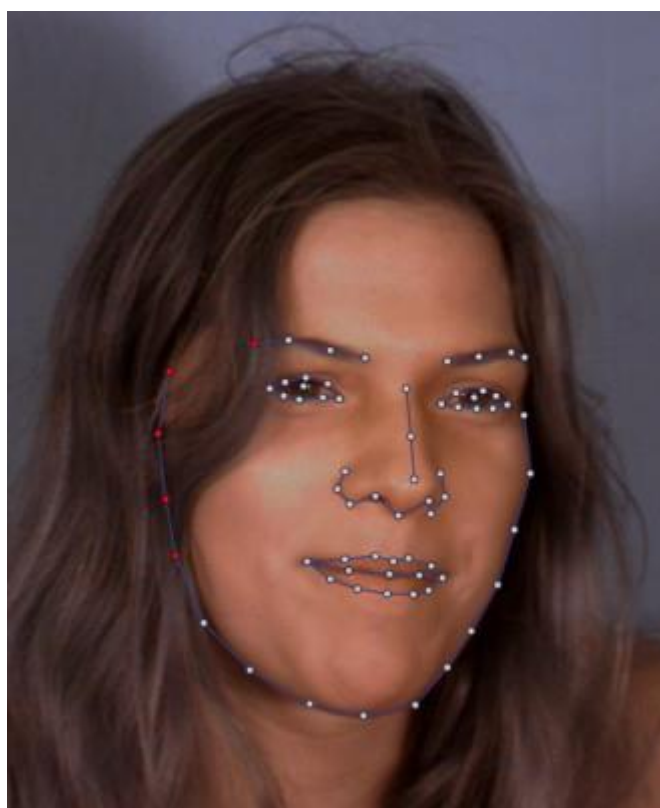


该模型具有一些用于分类的线性核。

4.2 面部模型

脸相对来说比较容易模仿，因为除了姿势和表情，每个人的脸并没有太大的差异。这样的模型可以手工构建，但是从带注释的数据中学习更容易，在我们的案例中，特征点被标记(注释)。由于注释面需要很长时间，所以我们使用了来自 MUCT 数据库的一些现有注释(稍微做了一些修改)，以及一些我们手工注释的面。

带注释的脸



来自突变数据库的带注释的面孔

为了从这些注释构建模型，我们使用主组件分析(简称 PCA)。我们首先计算所有标注的均值点，然后使用 PCA 将人脸的变化提取为向量或组件的线性组合。非常粗略地解释，PCA 将按重要性的顺序提取这些组件，即每个组件可以解释多少表面的变化。由于前几个部分能够覆盖面部姿势的大部分变化，所以我们可以不影响模型精度的情况下去掉其余部分。

PCA 提取的第一个成分通常包括基本的姿势变化，如偏航、俯仰，然后是张开和闭上的嘴、微笑等。





第一个提取的组件

然后，任何面部姿势都可以被建模为这些成分的平均值和加权组合，这些权重可以被认为是面部模型的“参数”。在这里查看完整的模型

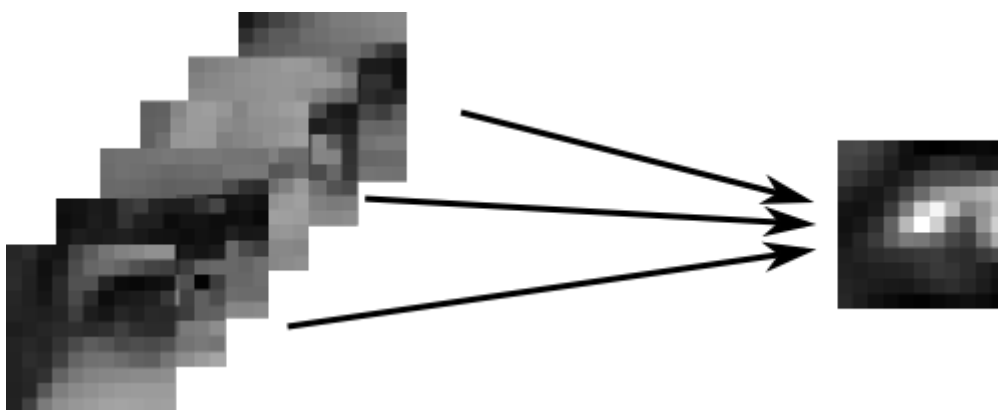
从 PCA 中，我们还存储了每个分量的特征值，它根据我们标注的数据[1]中的人脸姿态告诉我们每个分量的权重的标准差，这对于我们在优化步骤中规范化权重是非常有用的。

注意:PCA 不是提取参数化人脸模型的唯一方法。您还可以使用稀疏 PCA，它将导致“稀疏”转换。稀疏 PCA 在拟合/跟踪方面并没有给我们带来任何显著的改进，但通常会给我们提供看起来更自然的组件，这对于手动调整每个组件权重的正则化很有用。基于稀疏 PCA 的参数化人脸模型测试。

[1]:这也意味着用于训练模型的人脸是在各种不同的姿势和表情中很好的选择，否则我们最终会得到一个过于严格的规则化的模型，并且无法模拟“极端”的姿势

4.3 分类器

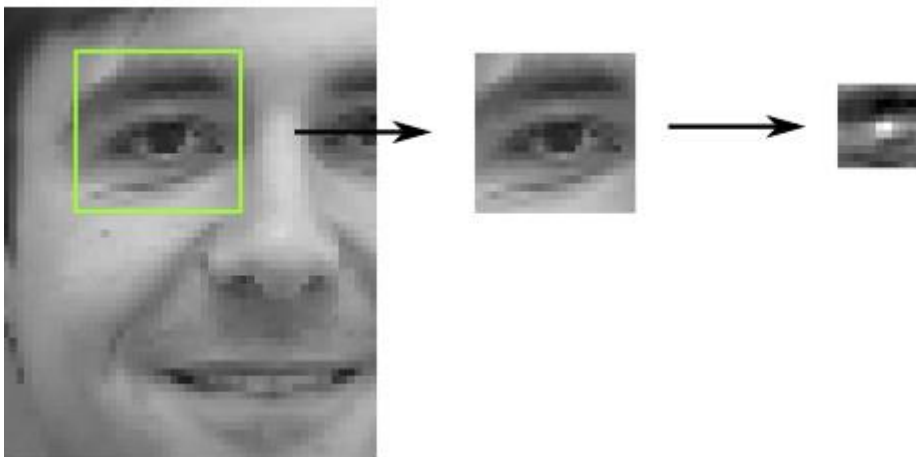
正如我所提到的，模型中的每个点都有一个分类器，因此我们的模型总共有 70 个分类器。为了训练这些分类器，例如第 27 点(左瞳孔)的分类器，我们在每个带注释的面部图像中以第 27 点的标记位置为中心裁剪一个 $X * X$ 的贴片。然后将这组补丁作为训练分类器的输入。



训练一个分类器，从带注释的面孔

我们使用的分类器可以是任何适合于图像分类的分类器，比如 Logistic 回归、SVM、正则相关过滤器，甚至是随机森林，但是在我们的例子中，我们实现了一个带有线性核的 SVM 分类器(这是原始论文建议的)，以及一个 MOSSE 过滤器。更多关于下面这些的实现问题。

当使用这些分类器来拟合模型时，我们在每个初始近似位置上裁剪一个搜索窗口，并在 searchwindow 中将各自的分类器应用到 $Y \times Y$ 像素的网格中。因此，我们得到一个 $Y * Y$ “响应” 输出，它将每个像素的概率映射为“对齐”的特征点。



剪短的补丁和左瞳孔的“响应”

4.4 优化

那么，既然我们有了分类器的反应，我们该如何应用这些信息以最合适的方式适合面部模型呢？

对于每个响应，我们计算模型的移动方式，以便以最高的可能性到达该区域。这是通过平均位移(大致相当于梯度下降)来计算的。然后，我们通过将“新位置”限制在面部模型的坐标空间中，来调整这个动作。通过这种方式，我们确保模型的点不会以与模型整体不一致的方式移动。这个过程是迭代进行的，这意味着面部模型将逐渐收敛到最优拟合[2]。

[2]:这恰好是期望-最大化的情况，根据响应找到最佳运动是期望步骤，对模型的正则化是最大化步骤

4.5 初始化

需要注意的一点是，由于我们使用的 searchwindows 非常小，如果在这些 searchwindows[3]的 “reach” 之外，该模型无法适应一个面孔。因此，关键是我们要在离模型 “真实” 位置不远的地方初始化模型。为了做到这一点，我们首先使用人脸检测器来查找人脸的粗糙边界框，然后通过相关过滤器识别眼睛和鼻子的大致位置。然后我们使用 procrustes 分析粗略地将平均面部模型拟合到眼睛和鼻子的已发现位置，并以此作为模型的初始位置。



人脸模型的初始化与拟合

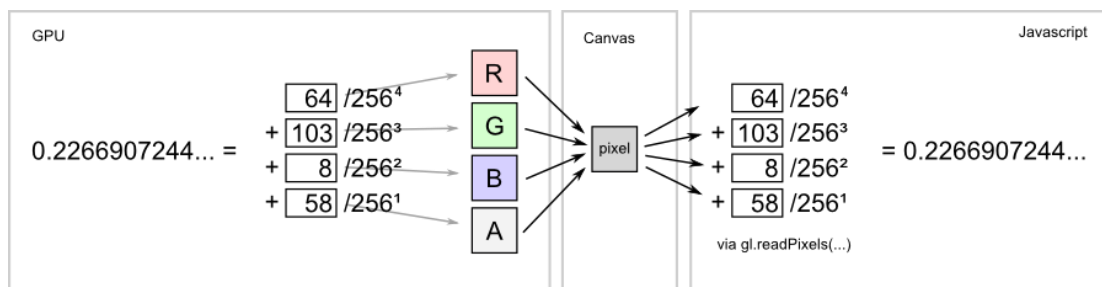
当我们在跟踪和装配脸部时，我们还需要检查模型是否偏离了脸的 “真实” 位置。一种方法是，每一秒钟检查一次，这样面部模型覆盖的近似区域，看起来就像一张脸。我们使用与贴片上相同的分类器，逻辑回归，只对整张脸进行训练。如果面部模型似乎不在人脸的顶部，我们将重新初始化人脸检测。

我们当然可以让 searchwindow 变得更大，但是我们每增加一个像素，我们的 searchwindow 就会增加适应时间，所以我们更喜欢使用小窗口

4.6 计算性能

这种算法在 javascript 中的直接实现非常缓慢。主要的瓶颈是在每次迭代中对模型中的每个点调用多次的分类器。根据 searchwindow (n)的大小和分类器补丁(m)的大小，直接实现是 $O(m^2 * n^2)$ 操作。通过 FFT 的卷积我们可以把它降低到 $O(n \log(n))$ 但这仍然比我们想要的要慢。幸运的是，线性内核非常适合通过 GPU 进行快速计算，我们可以通过 WebGL 实现这一点，目前大多数浏览器都有这种功能。当然，webGL 本来就不打算用于科学计算，它只用于图形化渲染，所以我们必须跨越一些障碍才能让它正常工作。

我们遇到的主要问题是，虽然大多数图形卡都支持浮点计算，而且我们可以很容易地将浮点数导入 GPU,但是无法在 WebGL 中将浮点数导出回 javascript。我们只能读取 GPU 渲染到画布上的像素(只支持 8 位的 ints)。要解决这个问题，我们必须使用一个技巧:我们将 32 位浮点数“打包”到 4 个 8 位的 int，通过将它们绘制到 canvas 来“导出”它们，然后读取这些像素，然后在 javascript 端将它们“解压缩”回 32 位的浮点数。在我们的例子中，我们将浮动分割为四个通道(R,G,B,A)，这意味着每个呈现的像素都包含一个浮动。尽管这对于一些性能调整来说似乎有很多麻烦，但这是值得的，因为 WebGL 实现的速度是 javascript 实现的两倍。



一旦我们得到了响应，我们就必须处理矩阵数学来进行正则化。这是另一个瓶颈，真正暴露了不同浏览器的 javascript 引擎之间的计算速度的巨大差异。我使用了优秀的“数字”库。做这些计算——它目前似乎是 javascript 中速度最快、功能最全面的矩阵库，我强烈推荐给任何想用 javascript 做矩阵运算的人。

在我们的最终基准测试中，我们设法在 Chrome 中每秒运行大约 70 次算法迭代(使用默认设置)，这足以实时地适应和跟踪人脸。

5.附件文献：基于正则化的地标均值偏移方法拟合变形模型

基于正则化的地标均值偏移方法拟合变形模型

Jason M. Saragih • Simon Lucey • Jeffrey F. Cohn

接收日期：2009 年 5 月 12 日/接受日期：2010 年 9 月 10 日/在线发布：2010 年 9 月 25 日
©Springer Science+Business Media, LLC 2010

摘要：

变形模型拟合在计算机视觉领域已经有十多年的历史了。因此，人们提出了许多方法，取得了不同程度的成功。一类已经显示出巨大希望的方法是，对模型地标的位置进行独立预测，通过对它们的关节运动强制执行一个先验值来组合。该方法创新的一个共同主题是用更简单的参数形式替换从每个局部检测器获得的可能地标位置的分布。在这项工作中，提出了一个原则性的优化策略，其中这些可能性的非参数表示在平滑的目标。结果更新方程-在地标上没有出现均值偏移，但在它们的关节运动上通过一个全局优先权进行调节。还提出了处理部分遮挡和降低计算复杂度的扩展。通过数值实验表明，该方法在一般人脸拟合任务上优于现有的一些常用方法。

关键词：可变形•配准•均值偏移

1 简介

将图像与可变形模型的匹配问题对应起来,这就意味着一个可参数化的物体在模型上的匹配问题。这是一个困难的问题,因为它涉及到高维的优化,由于光照条件、图像噪声、分辨率和内在的可变性,物体的外观在不同的实例之间会有很大的变化。针对这个问题,人们提出了许多方法,取得了不同程度的成功。其中,最有希望的一种方法是利用局部空间相干图像观测(即图像块)对物体进行建模(Cootes 和 Taylor 1992; Cristinacce 和 Cootes 2004、2006、2007; Wang 等人。2008a)。出于计算和泛化的目的,这些图像块被认为是有条件地相互独立的,这一假设在与近期文献中的整体方法比较时显示出了优越的性能(Liu 2007; Matthews and Baker 2004; Nguyen and De la Torre Frade 2008; Zhou and Comaniciu 2007)。局部图像块检测器通常是从标记的训练图像中学习对象中每个地标的。然而,由于其局部支持度小、训练时的外观变化大,这些局部检测器存在模糊性问题。这种模糊性可以在从每个地标探测器获得的地标位置(即响应图)的非参数分布中观察到。在这项工作中解决的中心难题是如何协同使用这些非参数措施,每个地标的可能位置,同时限制其模糊性的影响,当拟合可变形模型。

我们对解决这一困境的关键贡献在于认识到许多流行的优化策略在某种程度上都简化了从每个局部检测器参数化获得的地标位置的非参数分布。简化的目的是确保近似目标函数:

(i) 表现出使优化有效且数值稳定的特性,并且(ii)仍然大致保持与每个局部检测器相关的真实确定性/不确定性。接下来的问题是:为了满足(i)和(ii),应该如何简化这些局部分布?在这项工作中,我们对这个问题提出了一个新的答案,与迄今为止的方法不同,它不需要参数简化,但仍然确保了一个高效且数值稳定的优化。我们的非参数方法是已知和理解的均值漂移(Fukunaga and Hostetler 1975)模式搜索算法的再结晶。然而,这种方法不同于传统的均值漂移算法,因为它同时应用于所有的地标,并且对其联合运动施加了全局优先权。所得到的拟合算法简单、有效,并且在收敛速度和精度方面比现有方法有了显著的提高。

我们从第2节开始,详细介绍了可变形模型与有条件独立的路标检测拟合问题。许多方法在一个一致的公式下是统一的,其中可以更好地理解各种方法的观察行为。我们在第3节中介绍我们的方法,它利用了上一节中详述的公式。本节还介绍了对汉德尔部分遮挡的扩展,并降低了该方法的计算复杂度。通过实证实验,第4节将提出的方法与现有方法进行比较。我们在第5节结束时讨论并提到未来的工作。

2 背景

2.1 问题形成

大多数可变形模型拟合方法都采用线性近似来表示非刚性物体的形状如何变形,Cootes 和 Taylor (1992)提出了点分布模型(PDM)。它线性地建模非刚性形状变化,并将其与全局刚性变换组合,将形状放置在图像框中:

$$X_i = sR(\bar{x}_i + \Phi_{iq}) + t \quad (1)$$

其中,PDM的ITS地标和 $P = \{s, r, t, q\}$ 的二维位置表示PDM参数,其包括全局缩放 S 、旋转 R 、平移 T 和非刚性参数 Q 的集合。这里, x_i 表示 i^{th} PDM地标在参考帧中的平均位置。

(对于二维模型 i.e. $\bar{x}_i = [\bar{x}_i; \bar{y}_i]$), i 表示变化基的子矩阵,与那个地标有关。这样的模型既简单又有效,并且已经被证明能够充分地模拟物体的变形,例如人脸(Cootes and Taylor 1992)和医学图像分析中的器官(Zhou et al. 2005年)。

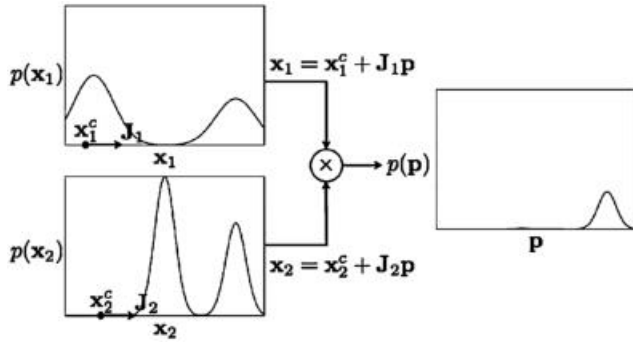


图 1 是由关节运动约束促进的检测模糊度降低的图示。地标 $\{x_1, x_2\}$ 受线性形式约束： $x_i = x_i^c + J_i p$ ，它们共同地将参数与其参数进行参数化。虽然独立的界标似然 $p(x_1)$ 和 $p(x_2)$ 是多模态的，但在联合运动约束下，模糊度被去除： $p(p) = p(x_1) p(x_2)$ 。

近年来，一种对所有 PDM 地标使用一组独立的局部检测器的方法(见 Cootes 和 Taylor 1992; Cristinacce 和 Cootes 2007、2004、2006; Wang 等人。2008a; 周等。例如，2005 年)已经吸引了一些兴趣，因为它避免了许多整体方法的缺点，比如建模复杂性和敏感性它能激发变化。影响在复杂的地标探测中，一个与假定的探测的有限支持区域直接相关的结果，通过限制这些地标关节运动的形状模型减少(见图 1)。虽然我们主要对利用统计形状模型的方法感兴趣，如(1)，但是这种框架的效用已经在光流等问题中使用了更一般的约束(Bruhn et al. 2005)和立体匹配(Sun 等人。2003 年)，约束以平滑过程的形式出现在地标的运动域上。在这项工作中，我们将这些方法统称为约束局部模型 (CLM)¹。

2.1.1 拟合目标

CLM 拟合通常被认为是对 PDM 参数 p 的搜索，该参数将所有地标上的偏差误差最小化，并进行适当的正则化：

$$Q(p) = R(p) + \sum_{i=1}^n D_i(x_i, I) \quad (2)$$

其中 R 惩罚复杂变形(即正则化项)， D_i 表示在图像 I (即数据项)中席上的 i^{th} 界标的不对中度量。正则化与描述合理物体形状的 PDM 参数的平均分布有关，常见的例子包括高斯分布(Basso et al. 2003)和高斯混合模型(GMM)(Gu 和 Kanade 2008)估计。错位误差函数的例子包括局部斑块外观上的马氏距离(Cootes 和 Taylor 1992)和基于增强 Harri-like 特征的分类器(Cristinacce 和 Cootes 2006)。虽然可以使用通用优化策略最小化(2)，但在实践中很少这样做。除了跟踪目标方法外， d_i 通常被选为模板和图像之间的最小平方差(Zhou 等人。2005 年)，大多数 CLM 公司的变体都采用了专门的融资策略。其中一个原因是，在席的空间域中，失调误差函数通常表现出明显的噪声。因此，局部确定性优化策略，如牛顿法，往往是不稳定的。随机优化策略，如 Cristinacce 和 Cootes (2004) 中使用的基于简单的方法，因为它们不使用梯度信息，这使得它们对测量噪声有些不敏感，因此更稳定。然而，当使用这些优化器时，收敛可能会很慢，特别是对于参数数目较多的 complexpdms。由于地标的未对准误差仅取决于其空间坐标，因此可以有效地执行对每个地标位置的独立彻底局部搜索(即在估计地标位置周围的所有整数像素位置)。因此，大多数 CLM 变体实施两步定位策略，首先执行彻底的本地搜索以获得每个地标的响应图。然后在这些响应图上执行优化，与不使用特定领域知识的通用优化方法相比，这些响应图允许使用更复杂的策略。这两个步骤的说明如图 2 所示。应该注意到，这是可能的限制搜索域{席} $n=1$ ，条件规格 C 到 CLM 的公式。第 2.2 节详细讨论了这些策略。

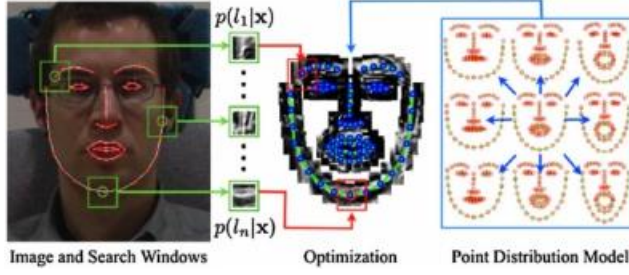


图 2 CLM 及其两个组成部分的说明：(i) 对特征位置进行彻底的局部搜索，以获得响应图 $\{p(l_i=1 | x, i)\}_{i=1}^n$ ，和 (ii) 优化策略，以最大化 PDM 约束地标的响应。

¹这一术语不应与 Cristinacce 和 Cootes (2006) 的工作混淆，后者是我们术语中 CLM 的一个特殊实例。

2.1.2 概率解释

(2) 中的 CLM 目标可以解释为使模型参数的可能性最大化，使其所有的地标与图像中对象上的相应位置对齐。目标的特定形式隐含地假设了探测到的目标和标记之间的条件独立性，其概率解释采用以下形式：

$$p(\mathbf{p}|\{l_i = 1\}_{i=1}^n, \mathcal{I}) \propto p(\mathbf{p}) \prod_{i=1}^n p(l_i = 1|\mathbf{x}_i, \mathcal{I}), \quad (3)$$

其中 $l_i \in \{1, -1\}$ 是一个离散随机变量，表示第 i 个地标是对齐还是未对齐。根据这个公式，

(2) 中的调节和失调误差函数的形式如下：

$$\mathcal{R}(\mathbf{p}) = -\ln\{p(\mathbf{p})\} \quad (4)$$

$$\mathcal{D}_i(\mathbf{x}_i; \mathcal{I}) = -\ln\{p(l_i = 1|\mathbf{x}_i, \mathcal{I})\}. \quad (5)$$

为了澄清以下章节中的论述，让我们详细说明 (3) 中用于本工作的先验和可能性的具体形式。我们对特定地标位置 \mathbf{x} 的线形可能性进行建模，如下所示：

$$p(l_i = 1|\mathbf{x}, \mathcal{I}) = \frac{1}{1 + \exp\{l_i c_i(\mathbf{x}; \mathcal{I})\}}, \quad (6)$$

其中 c_i 表示区分对齐位置和未对齐位置的分类器。注意，这种可能性是一个适当的概率质量函数，因为它在任何地方都是非负的，并且：

$$p(l_i = 1|\mathbf{x}, \mathcal{I}) + p(l_i = -1|\mathbf{x}, \mathcal{I}) = 1. \quad (7)$$

对于分类 c_i ，我们使用逻辑回归 (Wangetal. 2008a)：

$$c_i(\mathbf{x}; \mathcal{I}) = \mathbf{w}_i^T \mathcal{P}(\mathcal{W}(\mathbf{x}; \mathcal{I})) + b_i, \quad (8)$$

其中 $\{\mathbf{w}_i, b_i\}$ 分别表示增益和偏差， $\mathcal{P}(c)$ 将 c 归一化为零均值和单位方差。这里， $\mathcal{W}(\mathbf{x}; \mathcal{I})$ 是一个图像修补程序：

$$\mathcal{W}(\mathbf{x}; \mathcal{I}) = [\mathcal{I}(\mathbf{z}_1); \dots; \mathcal{I}(\mathbf{z}_P)]; \quad \{\mathbf{z}_i\}_{i=1}^P \in \Omega_{\mathbf{x}}, \quad (9)$$

其中 $\Omega_{\mathbf{x}}$ 表示以 \mathbf{x} 为中心的边界框内的整数像素位置集。使用该分类器的优点是可以使用有

效的卷积运算计算响应图。当假设 PDM 参数的先验信息是非信息的（一致的），则（3）中的公式将导致最大似然（ML）估计，否则将导致最大 a-后验（MAP）估计。当使用通过将 PCA 应用于一组注册形状获得的线性形状模型时，通常假定非刚性形状参数呈现高斯分布，导致以下先验：

$$p(\mathbf{p}) \propto \mathcal{N}(\mathbf{q}; \mathbf{0}, \Lambda); \quad \Lambda = \text{diag}\{[\lambda_1; \dots; \lambda_m]\}, \quad (10)$$

式中， λ_i 表示第 i 个非刚性变形模式的特征值。最后，一个非信息先验通常放在刚性变换上，该变换将模型放置在图像框中，这假设所有刚性变换的可能性相等。

2.2 现有装配策略

优化有两个难点：（1）如何避免局部最优，同时提供有效的评估；（2）如何处理边远检测。在下面的部分中，我们将展示现有的优化策略需要用更简单的参数形式替换真实的响应映射，并在这些参数上执行优化，而不是原始的响应映射。这些方法的相对性能可以通过在其近似值中对参数形式所作的特殊选择来解释。一般来说，近似响应图的复杂性决定了优化的计算成本和对局部极小值的敏感性，以及它对真实目标的表示的可信度。

2.2.1 各向同性高斯估计

最简单的 CLM 优化策略是在主动形状模型(ASM)中使用的策略，这是 Cootes 和 Taylor(1992) 首次提出的。该方法首先需要找到每个响应图中达到最大值的位置： $\mu = [\mu_1; \dots; \mu_n]$ 。优化程序的目标是使 PDM 和峰值响应坐标之间的加权最小二乘差最小化，并适当调整：

$$Q_{\text{ISO}}(\mathbf{p}) = \|\mathbf{q}\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n w_i \|\mathbf{x}_i - \mu_i\|^2, \quad (11)$$

其中权重 $\{w_i\}_{i=1}^n$ 反映了峰值响应坐标上的置信度，并且通常设置为响应 $\{\mu_i\}_{i=1}^n$ 的某个函数，使其更具有抵抗力

比如部分遮挡，被遮挡的路标的权重会更弱。通过采用 PDM 地标的一阶泰勒展开迭代最小化方程（11）：

$$\mathbf{x}_i \approx \mathbf{x}_i^c + \mathbf{J}_i \Delta \mathbf{p}, \quad (12)$$

解决参数更新问题：

$$\Delta \mathbf{p} = -\mathbf{H}_{\text{ISO}}^{-1} \left(\tilde{\Lambda}^{-1} \mathbf{p} + \sum_{i=1}^n w_i \mathbf{J}_i^T (\mathbf{x}_i^c - \mu_i) \right), \quad (13)$$

然后将其与电流参数相加： $\mathbf{p} \leftarrow \mathbf{p} + \Delta \mathbf{p}$ 。这里， $\tilde{\Lambda} = \text{diag}\{[0; \lambda_1; \dots; \lambda_m]\}$ ， $\mathbf{J} = [\mathbf{J}_1; \dots; \mathbf{J}_n]$ 是 PDM 的雅可比， $\mathbf{x}_c = [\mathbf{x}_{c1}; \dots; \mathbf{x}_{cn}]$ 是当前形状估计值，并且：

$$\mathbf{H}_{\text{ISO}} = \tilde{\Lambda}^{-1} + \sum_{i=1}^n w_i \mathbf{J}_i^T \mathbf{J}_i \quad (14)$$

是高斯牛顿黑森。从概率的角度来看，该算法的优化过程相当于用各向同性高斯估计量来近似响应映射²

² $p(l_i = 1 | \mathbf{x}_i, \mathcal{I})$ 和 $p(\mathbf{x}_i | l_i = 1, \mathcal{I})$ 的比例由 $p(\mathbf{x}_i | \mathcal{I})$ 是非信息的假设引起，这是假设所有刚性变换（在图像帧中放置形状）的直接结果同样可能。

$$p(l_i = 1 | \mathbf{x}_i, \mathcal{I}) \propto p(\mathbf{x}_i | l_i = 1, \mathcal{I}) \approx \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_i, \sigma_i^2 \mathbf{I}), \quad (15)$$

式中 $w_i = \sigma_i^{-2}$ 。很容易证明，将该近似值代入 (3) 并取其负对数结果，即可得到 (11) 中的目标值。

2.2.2 各向异性高斯估计

尽管上述近似值简单有效，但在某些情况下，它可能是对真实响应图的错误估计。首先，地标检测器，如 (8) 中的 logistic 回归，通常是不完美的，因为响应的最大值可能并不总是与 cor 一致。第二，矩形地标定位，因为检测中使用的特征是由小图像块组成的，它们通常包含有限的结构，从而导致检测模糊。一个常见的例子是孔径问题，在这个问题中，通过边缘的检测比沿着边缘的检测可信度更好（参见图 3 中鼻梁和下巴的响应图示例）。为了解释这些问题，许多作者建议将方向不确定性纳入响应图估计（见 Nickels 和 Hutchinson 2002; Zhou 等, 2005; Wang 等人, 例如 2008a）。与 (15) 中的近似值类似，这里的响应图由全协方差高斯分布近似：

$$p(l_i = 1 | \mathbf{x}_i, \mathcal{I}) \approx \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (16)$$

使用这种近似的各种方法之间的主要区别在于如何估计平均值和协方差。在 Nickels 和 Hutchinson (2002) 中， $\boldsymbol{\mu}_i$ 被选为真实响应图中的最大值，协方差设置为 ML 溶液：

$$\boldsymbol{\Sigma}_i = \sum_{\mathbf{x} \in \Psi_i} \frac{p(l_i = 1 | \mathbf{x}, \mathcal{I})}{\sum_{\mathbf{y} \in \Psi_i} p(l_i = 1 | \mathbf{y}, \mathcal{I})} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^T, \quad (17)$$

其中 Ψ_i 是 2d 矩形网格，在该网格上执行强制局部搜索（即搜索窗口）。在 Wang 等人 (2008a) 中，凸二次函数是响应图的负对数，从中可以推断近似密度的平均值和协方差。In Zhou et al. (2005)，其中，将总平方差用作地标 fit 的测量，拉普拉斯近似 (Gelman 等人, 1995) 与小运动近似结合使用，得出标度文法作为协方差估计，平均值定义为 ML 光学流解，标度定义为该解的外观误差方差。无论在计算响应图的各向异性高斯估计时使用何种策略，通过将此近似代入 (3) 中的目标，优化问题可以写成以下最小值：

$$\mathcal{Q}_{\text{ANI}}(\mathbf{p}) = \|\mathbf{q}\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\mu}_i\|_{\boldsymbol{\Sigma}_i^{-1}}^2, \quad (18)$$

可以迭代求解。此目标的高斯-牛顿更新采用以下形式：

$$\Delta \mathbf{p} = -\mathbf{H}_{\text{ANI}}^{-1} \left(\tilde{\Lambda}^{-1} \mathbf{p} + \sum_{i=1}^n \mathbf{J}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_i^c - \boldsymbol{\mu}_i) \right), \quad (19)$$

其中

$$\mathbf{H}_{\text{ANI}} = \tilde{\Lambda}^{-1} + \sum_{i=1}^n \mathbf{J}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{J}_i. \quad (20)$$

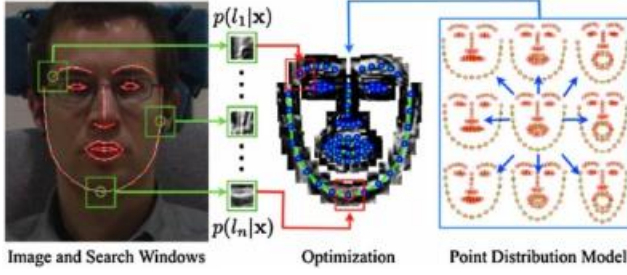


图 3 不同方法中使用的响应图、RES 和它们的近似值，适用于最外层的角落，然后是桥和下巴。交叉在响应图上表示真正的地标位置。GMM 近似有五个簇中心。给出了 $\rho \in \{20, 5, 1\}$ 的 KDE 近似。

2.2.3 高斯混合模型估计

虽然响应图的各向异性高斯近似可以克服其各向同性对应的一些缺点，但在某些情况下，其估计过程可能很差。特别是，当响应图是强多模的时，这种近似会对各种模式进行平滑处理（参见图 3 中眼角的示例响应图），从而限制了最终结果的精确性。为此，在 Gu 和 Kanade (2008) 中，使用高斯混合模型 (GMM) 来近似响应图：

$$p(l_i = 1 | \mathbf{x}_i, \mathcal{I}) \approx \sum_{k=1}^{K_i} \pi_{ik} \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}), \quad (21)$$

式中， K_i 表示模式数， $\{\pi_{ik}\}_{k=1}^{K_i}$ 是 i^{th} 地标 GMM 的混合系数。将每个地标的模式成员 $\{z_i\}_{i=1}^n$ 视为隐藏变量，可以使用期望最大化 (EM) 算法找到最大似然解，该算法最大化：

$$p(\mathbf{p} | \{l_i\}_{i=1}^n, \mathcal{I}) \propto p(\mathbf{p}) \prod_{i=1}^n \sum_{k=1}^{K_i} p_i(z_i = k, l_i | \mathbf{x}_i, \mathcal{I}) \quad (22)$$

对于 $\{l_i = 1\}_{i=1}^n$ 。EM 算法的 E 步骤涉及计算潜在变量 $\{z_i\}_{i=1}^n$ 的后验分布：

$$p(z_i = k | l_i, \mathbf{x}_i, \mathcal{I}) = \frac{p(z_i = k) p(l_i | z_i = k, \mathbf{x}_i, \mathcal{I})}{\sum_{j=1}^{K_i} p(z_i = j) p(l_i | z_i = j, \mathbf{x}_i, \mathcal{I})}, \quad (23)$$

其中 $\{l_i = 1\}_{i=1}^n$, $p(z_i = k) = \pi_{ik}$ ，并且

$$p(l_i = 1 | z_i = k, \mathbf{x}_i, \mathcal{I}) = \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}). \quad (24)$$

让 $q(\mathbf{z}) = \prod_{i=1}^n p_i(z_i | l_i = 1, \mathbf{x}_i, \mathcal{I})$ ，EM 算法的 M 步包括最小化完整数据的负对数的期望：

$$\begin{aligned} \mathcal{Q}_{\text{GMM}}(\mathbf{p}) &= E_{q(\mathbf{z})} \left[-\ln \left\{ p(\mathbf{p}) \prod_{i=1}^n p(l_i = 1, z_i | \mathbf{x}_i, \mathcal{I}) \right\} \right] \\ &\propto \|\mathbf{q}\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n \sum_{k=1}^{K_i} w_{ik} \|\mathbf{x}_i - \boldsymbol{\mu}_{ik}\|_{\boldsymbol{\Sigma}_{ik}^{-1}}^2, \quad (25) \end{aligned}$$

其中 $w_{ik} = P(Z_i = K = L_i = 1, \text{席}, I)$ 。该目标可使用高斯-牛顿优化程序迭代最小化，更新形式如下：

$$\Delta \mathbf{p} = -\mathbf{H}_{\text{GMM}}^{-1} \left(\tilde{\Lambda}^{-1} \mathbf{p} + \sum_{i=1}^n \sum_{k=1}^{K_i} w_{ik} \mathbf{J}_i^T \Sigma_{ik}^{-1} \Delta \mathbf{x}_{ik} \right), \quad (26)$$

其中 $\Delta \mathbf{x}_{ik} = \boldsymbol{\mu}_{ik} - \mathbf{x}_i^c$ ，并且

$$\mathbf{H}_{\text{GMM}} = \tilde{\Lambda}^{-1} + \sum_{i=1}^n \sum_{k=1}^{K_i} w_{ik} \mathbf{J}_i^T \Sigma_{ik}^{-1} \mathbf{J}_i. \quad (27)$$

注意，高斯-牛顿优化程序不能保证非线性最小二乘函数的收敛到全局最小值（对于非线性最小二乘函数，由于姿态和非刚性参数在形状模型中的双线性关系，这是一个全局最小值）。因此，这一策略影响了广义 EM 算法（Dempster 等人，1977 年）。

虽然 GMM 比高斯近似更接近于响应图，但它有两个主要缺点。首先，由响应图估计 GMM 参数的过程本身就是一个非线性优化问题。它只是局部收敛的，并且需要预先选择模式的数目。由于每个 PDM 地标都需要 gm fitting，因此它构成了一个巨大的计算开销。虽然可以进行一些近似，但它们通常是次优的。例如，在 Gu 和 Kanade (2008) 中，模式被选为地图中最大的 K_i 响应。协方差是各向同性参数化的，其方差试探性地设置为拟合算法上一次迭代中到最近模式的缩放距离。这种近似可以有效地估计 GMM 参数，而不需要对真实响应图进行低近似的复杂过程。GMM 响应图近似的第二个缺点是 (22) 中的近似目标几乎总是多峰的。因此，采用 GMM 简化方法很容易在局部最优中终止。

3 正则化地标均值偏移

由于主成分分析中的截断，形状模型不能很好地重建真实的地标位置图像。误差估计从观测噪声出发，常被建模为同方差各向同性高斯：

$$\mathbf{y}_i = \mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad \text{where } \boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{\epsilon}_i; \mathbf{0}, \rho \mathbf{I}). \quad (28)$$

这里， ρ 表示地标位置上噪声的方差，可从以下训练集中推断出来（Moghadam 和 Pentland 1997）：

$$\rho = \frac{1}{N - m} \sum_{i=m+1}^N \lambda_i, \quad (29)$$

它是正交于 Φ 的子空间中特征值的算术平均值。让我们假设这里有一个定义 $\{\Psi_i\}_{i=1}^n$ 的模型的每个陆块的位置和日期。这可能是由于，例如当架构约束到一个局部矩形区域时，其中 Ψ_i 表示该区域内的所有整数像素位置。将实际地标的位置视为不可变量，我们将其边缘化，排除地标对齐的可能性：

$$p(l_i = 1 | \mathbf{x}_i, \mathcal{I}) = \sum_{\mathbf{y}_i \in \Psi_i} p(l_i = 1 | \mathbf{y}_i, \mathcal{I}) p(\mathbf{y}_i | \mathbf{x}_i), \quad (30)$$

其中，从 (28) 中，我们有：

$$p(\mathbf{y}_i | \mathbf{x}_i) = \mathcal{N}(\mathbf{y}_i; \mathbf{x}_i, \rho \mathbf{I}). \quad (31)$$

在 (30) 中, $p(l_i=1 | \mathbf{y}_i, \mathcal{I})$ 表示第 i 个地标与图 1 中的位置 \mathbf{y}_i 对齐的可能性。这里的公式与第 2.2 节中讨论的现有安装策略之间的主要区别。2.2 是响应图只在 \mathbf{Y} 位置通过 \mathbf{y}_i 进行评价, 而在 \mathbf{p} 上进行优化, 仅影响 $\{\mathbf{y}_i\}_{i=1}^n$ 。因此, 与其用单粒子参数形式近似响应 $m_{\mathbf{p}_i}$, 不如用同构各向同性高斯核密度估计 (KDE)³ (Silverman 1986):

$$p(l_i = 1 | \mathbf{x}_i, \mathcal{I}) = \sum_{\mathbf{y}_i \in \Psi_i} \pi_{\mathbf{y}_i} \mathcal{N}(\mathbf{x}_i; \mathbf{y}_i, \rho \mathbf{I}), \quad (32)$$

其中 $\pi_{\mathbf{y}_i} = p(l_i = 1 | \mathbf{y}_i, \mathcal{I})$ 。

响应的这种非参数估计的质量在选择和日期地标位置集 $\{\Psi_i\}_{i=1}^n$ 。如果候选样本太少, 它们可能无法充分覆盖变化空间, 并且从 (29) 中的训练数据中获得的 ρ 将被低估。然而, 由于 \mathbf{y}_i 的空间是 2D 图像平面, 因此通过对所有整数像素位置进行彻底的局部搜索, 可以在当前 PDM 估计值周围局部计算密集候选集的可能性。将 (32) 代入 (3), 得到:

$$p(\mathbf{p} | \{l_i = 1\}_{i=1}^n, \mathcal{I}) \propto p(\mathbf{p}) \prod_{i=1}^n \sum_{\mathbf{y}_i \in \Psi_i} \pi_{\mathbf{y}_i} \mathcal{N}(\mathbf{x}_i; \mathbf{y}_i, \rho \mathbf{I}). \quad (33)$$

值得注意的是, 使用信念传播 (Yedidia et al. 2002) 的结果与 PDM 参数的边际似然完全相同。关于这种观点的阐述可以在附录中找到。方程 (33) 可通过最大化法最大化, 如第 2.2 节所述的近似方法。2.2 条。处理真正的地标位置 $\{\mathbf{y}_i\}_{i=1}^n$ 作为隐藏变量, 在 E-step 中对候选对象的后验进行评估:

$$w_{\mathbf{y}_i} = p(\mathbf{y}_i | l_i = 1, \mathbf{x}_i, \mathcal{I}) = \frac{\pi_{\mathbf{y}_i} \mathcal{N}(\mathbf{x}_i; \mathbf{y}_i, \rho \mathbf{I})}{\sum_{\mathbf{z}_i \in \Psi_i} \pi_{\mathbf{z}_i} \mathcal{N}(\mathbf{x}_i; \mathbf{z}_i, \rho \mathbf{I})}. \quad (34)$$

然后, M 步骤包括最小化:

$$\begin{aligned} Q_{\text{KDE}}(\mathbf{p}) &= E_{q(\mathbf{y})} \left[-\ln \left\{ p(\mathbf{p}) \prod_{i=1}^n p(l_i = 1, \mathbf{y}_i | \mathbf{x}_i, \mathcal{I}) \right\} \right] \\ &\propto \|\mathbf{p}\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n \sum_{\mathbf{y}_i \in \Psi_i} \frac{w_{\mathbf{y}_i}}{\rho} \|\mathbf{x}_i - \mathbf{y}_i\|^2, \end{aligned} \quad (35)$$

³ 由于优化超过模型参数 \mathbf{p} , 它仅影响 $\{\mathbf{x}_i\}_{i=1}^n$, 而不是其席候选 $\{\mathbf{y}_i\}_{i=1}^n$, 我们已经取代了 (31) 中的高斯分布的均值和变量。

其中 $q(\mathbf{y}) = \prod_{i=1}^n p(\mathbf{y}_i | l_i = 1, \mathbf{x}_i, \mathcal{I})$ 利用关系式: $\sum_{\mathbf{y}_i \in \Psi_i} w_{\mathbf{y}_i} = 1$, 线性化形状模型的解可以写成:

$$\Delta \mathbf{p} = -(\rho \tilde{\Lambda}^{-1} + \mathbf{J}^T \mathbf{J})^{-1} (\rho \tilde{\Lambda}^{-1} \mathbf{p} - \mathbf{J}^T \mathbf{v}), \quad (36)$$

其中 $\mathbf{v} = [v_1; \dots; v_n]$ 是每个地标的平均偏移向量的串联:

$$\mathbf{v}_i = \left(\sum_{\mathbf{y}_i \in \Psi_i} \frac{\pi_{\mathbf{y}_i} \mathcal{N}(\mathbf{x}_i^c; \mathbf{y}_i, \rho \mathbf{I})}{\sum_{\mathbf{z}_i \in \Psi_i} \pi_{\mathbf{z}_i} \mathcal{N}(\mathbf{x}_i^c; \mathbf{z}_i, \rho \mathbf{I})} \mathbf{y}_i \right) - \mathbf{x}_i^c. \quad (37)$$

注意, 在需要 ML 解⁴ 的情况下 (即 $p(\mathbf{p})$ 是非信息的), 参数更新 $\Delta \mathbf{p}$ 的解决方案仅仅是均值偏移向量到由 PDM 的 Jacobian 跨越的子空间上的非正交投影。在任何情况下, 在 (36)

中的解决方案建议了一个简单而有效的实现,包括计算均值偏移向量和通过形状模型对其进行正则化之间的交替。算法 1 概述了完整的拟合过程,我们将其称为正则化地标均值偏移 (RLMS)。Fashing 和 Tomasi(2005)曾指出均值漂移是一个有界优化。这一点后来在 Carreira Perpinan (2007) 中得到了扩展,表明对于高斯核,均值漂移相当于采用 EM 算法作为优化策略。通过对 RLMS 的推导,我们证明了这种解释可以进一步推广到具有条件独立可能性的问题。因此,RLMS 优化策略利用了 EM 算法的可证明收敛性和改进性。

Algorithm 1 Regularized landmark mean-shift

Require: \mathcal{I} and \mathbf{p}

- 1: Compute responses $\{(6)\}$
 - 2: **while** not_converged(\mathbf{p}) **do**
 - 3: Linearize shape model $\{(12)\}$
 - 4: Compute mean-shift vectors $\{(37)\}$
 - 5: Compute PDM parameter update $\{(36)\}$
 - 6: Update parameters: $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$
 - 7: **end while**
 - 8: **return** \mathbf{p}
-

3.1 避免局部最小值

第节中讨论的响应图近似值。2.2 可以看作是一种平滑的形式。这解释了各种方法的相对性能。高斯近似最平滑,但近似真实响应图最差,而 GMMs 影响的平滑不那么积极,但对局部最优表现出一定程度的敏感性。我们可以考虑同时使用高斯和 GMM 近似,其中高斯近似用于得到 GMM 近似的收敛域。然而,这样一种高斯近似的方法不能保证高斯近似的模式位于 GMM 的收敛域内。利用 RLMS 中的 KDE 近似,可以设计出一种更优雅的方法,其中,响应图估计的复杂性直接由高斯核的方差控制(见图 3)。这里的指导原则类似于在高斯金字塔上进行优化。可以看出,当使用高斯核函数时,存在 $\rho < \infty$,使得 KDE 是单峰的,而不

⁴本节中的推导是我们最初在 Saragih 等人中提出的 ML 公式的扩展。(2009 年)。

考虑样本的分布 (Carreira Perpinan and Williams 2003)。当 ρ 减小时,目标地形的模式区分度和平滑度降低。然而,在较大的 ρ 处物镜的最优值很可能最接近具有较小 ρ 的物镜的期望模式,从而促进其收敛到正确的模式。因此, ρ 减小的策略将引导优化朝着真正目标的全局最优方向发展。应该注意的是,在制定 RLMS 中,选择一个具有里程碑意义的候选的 Geang-KalnIsIn 因子的概念,即 P-的里程碑估计: $P(Y_i \times \text{席})$ 。因此,给定局部搜索中足够的粒度(即选择 $\{i\}_{n_i=1}$),那么最能代表类似情况的核方差 ρ 。

3.2 处理部分闭塞

其中一个主要的模仿和现存的策略描述在第节。2.2 是标记的可能性是一致的, $P(L_i=1 \times \text{席 } I, I)$,不编码关于闭塞的影响的信息。通常,它是在一组从对齐(在区分分类器的情况下是未对齐的)位置 i 裁剪的图像块上训练的没有无障碍图像。由于不可能完全覆盖整个空间,因此在大多数流媒体应用程序中,为被遮挡的地标建模的对齐方式是不可操作的被遮挡的外观(即被遮挡的物体可以然而,当图像中感兴趣的对象被部分遮挡时,在训练过程中忽略这些情况可能会产生有害的影响。

不符合假设模型的观测值通常称为异常值。在 CLM 中,异常值来源于非高斯图像噪声、看不

见的外观和遮挡。在医学图像分析中经常观察到的非高斯图像噪声的一个例子是信号丢失效应 (Zhou et al. 2005 年)。虽然在整个图像上是非高斯的, 但是这些异常值可以通过现有的拟合策略得到很好的处理, 因为它们的主要作用是增加似然的空间分布, 这相当于增加了其估计的不确定性。这反过来导致对全局目标的贡献较小, 限制恶化不受噪音影响。异常值从看不见的外观衍生出来在克隆框架中没有那么大的问题, 因为在学习对齐可能性时只考虑补丁的外观。这些补丁通常非常小 (即, 在 (11×11) 像素的顺序)。因此, 对于许多问题, 存在足够的数据量来训练一个模型, 该模型可以很好地概括对象的大多数实例。由于光照变化引起的变化可以部分地通过功率归一化来解释 (Wang 等人, 2008a)。这与基于整体的方法形成了鲜明的对比 (例如, 参见 Liu 2007; Matthews 和 Baker 2004; Nguyen 和 De la Torre Frade 2008; Zhou 和 Comaniciu 2007), 后者考虑了对象内所有像素之间的相关性。

虽然现有的方法对上述两种异常值类型有一定的鲁棒性, 但对于遮挡情况则不能这样说。原因是遮挡对象的面片外观可能与感兴趣对象的面片外观相似, 这是对齐的 Landmark 所使用的 ConditionalDependency 的直接结果可能性。例如, 在面部边缘周围出现的地标, 很容易与任何遮挡混淆具有强边的物体。

(36) 中的参数更新内部优化策略本质上是将每个地标的平均偏移向量正则化投影到合理形状变化的空间上。如前所述, 由于在遮挡存在时具有误导性的地标可能性, 被遮挡地标的均值偏移向量可能是错误的。因此, 用最小二乘投影正则化解不再合适。因此, 处理这种情况的一个简单方法是对这个投影使用 M 估计量。从形式上讲, 这需要用以下内容代替 (35) 中的 Q 函数:

$$Q_{\text{KDE}}(\mathbf{p}) \propto \|\mathbf{q}\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n \sum_{\mathbf{y}_i \in \Psi_i} \omega_{\mathbf{y}_i} \varrho(\|\mathbf{x}_i - \mathbf{y}_i\|^2; \boldsymbol{\theta}), \quad (38)$$

其中 ϱ 是 NM 估计器, 例如 Geman-McClure 函数, 该函数已广泛用于光流估计 (见 Black 和 Anandan 1993; Blake 等人, 例如 1994 年):

$$\varrho(r^2; \boldsymbol{\theta}) = \frac{r^2}{r^2 + \alpha^2}; \quad \boldsymbol{\theta} = \{\alpha\}. \quad (39)$$

然后, 方程 (38) 可通过使用迭代重加权最小二乘法进行求解, 如下所示:

$$\Delta \mathbf{p} = -\mathbf{H}_{\text{KDE}}^{-1} \left[\tilde{\Lambda}^{-1} \mathbf{p} + \sum_{i=1}^n \mathbf{J}_i^T \sum_{\mathbf{y}_i \in \Psi_i} \omega_{\mathbf{y}_i} \varrho'(\mathbf{x}_i^c - \mathbf{y}_i) \right], \quad (40)$$

其中 ϱ' 表示在 $\|\mathbf{x}_i^c - \mathbf{y}_i\|^2$ 处评估的估计量的导数, Hessian 的形式如下:

$$\mathbf{H}_{\text{KDE}} = \tilde{\Lambda}^{-1} + \sum_{i=1}^n \left(\sum_{\mathbf{y}_i \in \Psi_i} \omega_{\mathbf{y}_i} \varrho' \right) \mathbf{J}_i^T \mathbf{J}_i. \quad (41)$$

由于 RLMS 的优化策略本质上是广义 EM 算法, 为了保持 EM 优化的性质, 即收敛性和可证明性改进, 必须调整后验权重 ω_i , 以重新根据 ϱ (i.e. $\omega_{\mathbf{y}_i} \neq w_{\mathbf{y}_i}$ from (34)) 的选择来重新确定 P (Yi x 席) 的新参数化。对于特定的, 里程碑式候选可能性的形式如下:

$$p(\mathbf{y}_i|\mathbf{x}_i) = \frac{1}{Z(\boldsymbol{\theta})} \exp\{-\varrho(\|\mathbf{x}_i - \mathbf{y}_i\|^2; \boldsymbol{\theta})\}, \quad (42)$$

其中 Z 是一个分区（规范化）函数，它将（42）强制为 PDF：

$$Z(\boldsymbol{\theta}) = \int_{\mathbf{r}_i} p(\mathbf{r}_i|\mathbf{x}_i) d\mathbf{r}_i < \infty; \quad \mathbf{r}_i \in \Omega, \quad (43)$$

用 Ω 表示图像的有界空间域。当选择相同的超参数 $\boldsymbol{\theta}$ 作为所有地标的候选的 M-估计量时（即同构 KDE），则不需要显式地计算配分函数，因为它作为标度常数从方程中剔除。候选人的后面可以写为：

$$\omega_{\mathbf{y}_i} = \frac{\pi_{\mathbf{y}_i} \exp\{-\varrho(\|\mathbf{x}_i - \mathbf{y}_i\|^2; \boldsymbol{\theta})\}}{\sum_{\mathbf{z}_i \in \Psi_i} \pi_{\mathbf{z}_i} \exp\{-\varrho(\|\mathbf{x}_i - \mathbf{z}_i\|^2; \boldsymbol{\theta})\}}. \quad (44)$$

通过使用鲁棒误差函数来减少边缘候选点的影响，得到的算法相当于非高斯核的 RLMS，其中核的类型取决于的选择。然而，这种选择带来的一个复杂因素是超参数 $\boldsymbol{\theta}$ 的选择。对于高斯核的情况： $\boldsymbol{\theta} = \{\rho\}$ ，其最佳设置在（29）中给出。对于更一般的，不存在最优 $\boldsymbol{\theta}$ 的闭式解。尽管如此，出于某些合理的考虑，可以不考虑蒙特卡罗战略进行估算。例如，对于（39）中的 Geman-McClure 鲁棒函数，内线区域由： $|\mathbf{r}| \leq \alpha \sqrt{3}$ 给出。由于假定参数形状模型，内部标准差由 $\sqrt{\rho}$ 给出。因此，以下（罗伯茨等人。2007）我们可以设置： $\alpha = \gamma \sqrt{3} \rho$ ，其中 γ 是内部标准差的倍数，通常取 $\gamma \in [1, 3]$ 。

最后，需要注意的是，本文所述的 RLM 的稳健公式定义了一些不具说服力的边远地标候选人进入形状模型。尽管这种假设在很多情况下都是合理的，当然，一个边远的候选者与形状模型是一致的，但并不代表真正的地标位置。随着形状模型变化方向的数量增加，这种情况越来越有可能发生。虽然由先验 $p(\rho)$ 诱导的正则化方法部分地解决了这个问题，但它并不是一个完整的解决方案，如何最好地处理这种情况仍然是一个悬而未决的问题。

3.3 实际考虑

3.3.1 相似性规范化搜索

由于用于填充的穷尽局部搜索仅在空间维度上执行，因此这种方法很难解释尺度和平面内旋转的显著变化。然而，根据显示这些变化的图像训练的地标探测器将缺乏特定性，限制了拟合程序的精度。

例如，生成整体可变形模型拟合算法（Edwards 等人。1998 年；Matthews 和 Baker 2004 年），通常以预先确定的参考形状测量模型 fit ，通常根据训练数据的平均形状进行选择。我们在这里使用相同的原理，即使用 PDM 的当前比例和旋转估计将图像转换为参考帧。算法 1 的拟合过程是对该相似性标准化图像执行。然后，通过组合用于规范化图像的相似性变换的收敛形状，可以找到描述图像帧中形状的 PDM 参数。随着拟合过程的收敛，相似度变换的估计值接近目标的真实姿态，提高了地标检测的可靠性。

3.3.2 一维与二维搜索区域

CLM 公式中剩下的一个开放性问题是为每个地标 $\{i\}_{n \ i=1}$ 选择搜索区域，沿着通常（手动）选择垂直于地标外观最大边缘方向的文件进行彻底的局部搜索。其他人在当前地标估计值周围的矩形边界框内搜索（Cristinacce 和 Cootes 2007；Wang 等人。2008a）。

进行专业搜索的动机是双重的。首先，它的计算成本要低得多，只需要对 d -探测器进行评估，而对于正方形搜索区域， d_2 则是一个比较好的选择。其次，它利用了用于检测的补丁

的有限结构。如第节所述。2.2, 位于边缘的地标, 边缘上的最小电容量较低。因此, 通过对这些位置的探测器进行评估, 可以获得很少的运动信息。

虽然许多可变形对象的地标的常见选择确实是放置在边上, 但这并不严格必需的。例如, 人脸上的地标几乎总是包括眼角和唇角, 因为它们有助于更好地保持手动标注的一致性。对于这些地标, 局部结构足以将其与相邻位置的图像块区分开。因此, 将搜索限制在一个文件可能会使估计的目标函数产生偏差。这一点由于在垂直于搜索文件的方向上的不确定性而经常做出的启发式选择而变得更糟 (即, 通常设置为等同于沿着搜索方向的不确定性, Cootes 和 Taylor 1992; Gu 和 Kanade 2008)。在确定这些地标建筑的最佳方向方面也存在一些困难。虽然可以通过交叉验证策略从数据中学习最佳方向, 但大多数方法只是将方向定义为相邻地标各自位置的函数。最后, 为了确保可靠的拟合, 应选择适当的方向, 使其充分覆盖方向空间 (即, 仅具有水平搜索特性的模型不能垂直移动)。

矩形搜索区域不受预测的影响少学。他们唯一的缺点是它们涉及到更多的探测器评估, 这可能导致在使用复杂探测器时效率低下。在这项工作中, 我们使用矩形搜索区域, 因为用于检测的线性分类器可以使用有效的卷积运算进行评估。当然可以将 profile 和矩形搜索区域结合起来, 其中搜索区域的选择是针对每个地标显示的局部结构进行的, 但是我们在这项工作中并不追求这种方法。

3.3.3 效率预计算网格

在响应映射的 KDE 表示中, 核心中心被放置在搜索窗口定义的网格节点上。从 GMM 的角度来看, 这些内核代表了真正地标位置的候选。虽然不需要优化来确定模式的数量、它们的中心和混合系数, 但这里使用的候选者数量远大于一般 GMM 估计中通常使用的数量 (即基于 GMM 的表示通常使用 $K_i < 10$, 而搜索窗口的特征值大于 100 个节点), 在 (34) 中计算后验概率将更为昂贵。但是, 如果方差 ρ 是先验已知的 (见第 3.1), 则可以进行一些近似以显著降低计算复杂性。计算 mean-shift 更新的主要开销是计算当前地标估计值和响应图中每个网格节点之间的内核。由于网格位置是固定的, 假定假定为 ρ , 可以选择为席的各种设置预先计算内核。特别地, 一个简单的选择是沿着以响应图网格 i 的分辨率或更高的分辨率采样的网格预计算这些值。在拟合过程中, 只需确定该网格中最接近 PDM 地标当前估计值的位置, 并通过假设地标实际位于该节点来估计核心评估值 (见图 4)。这涉及一个表查找, 可以有效地执行。网格的粒度越高, 近似值就越好, 以更大的存储需求为代价, 但不会显著增加计算复杂度。查找表上的插值过程可能会进一步提高估计精度, 但我们发现, 鉴于网格的足够粒度, 这样的添加几乎没有整体效益。最后, 尽管这种近似破坏了严格改进的 EM 属性, 我们在第节中以经验证明了这一点。4 这种近似值仍然可以实现精确拟合。在我们的实现中, 我们发现这样的近似值将平均安装时间缩短了一半。

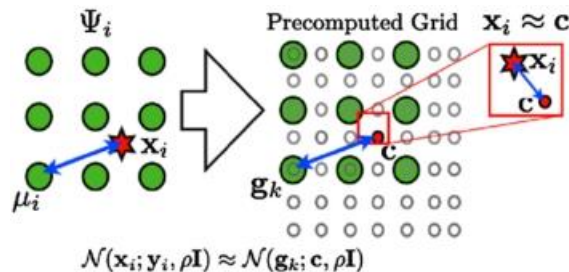


图 4 使用预计算网格进行有效均值偏移的说明。在 c 和网格中的所有其他节点之间预先计算内核计算。为了近似真实的核评估, 假设席与 C 一致, 任何响应图网格位置的可能性都可以通过查表来实现。

4 实验

4.1 拟合方法的实证比较

4.1.1 静止图像

以上讨论的各种 CLM 优化策略在两个公共可用数据库上的通用正面匹配问题上进行了比较：

(i) CMU 姿势、照明和表情数据库 (MultiPie) (Gross 等人, 2008 年), 以及 (ii) XM2VTS 数据库 (Messer 等人, 1999 年)。MultiPie 用 68 个点的标记进行注释, 用作基本事实地标。我们使用了 339 名受试者的 762 张正面面部图像。XM2VTS 由 295 个受试者的 2360 个正面面部图像组成, 这些图像的基本真实性注释是公开的, 但与我们对 MultiPie 的 68 点标记不同。XM2VTS 只包含中性表达式, 而 MultiPie 包含重要表达式变化。多个物体的平均中心距离 XM2VTS 数据库分别为 80 和 100 像素。在 MultiPie 和 XM2VTS 上分别进行了四次交叉验证实验, 将图像分成四组不重叠的被试身份。在每个试验中, 三个分区用于训练, 其余部分用于测试。

在这些数据库中, 我们比较了四种类型的优化策略: (i) ASM (Cootes and Taylor 1992) 中描述的。2.2.1, (ii) 凸二次拟合 (CQF) (Wang 等人, 2008a) 在第 2.2.2, (iii) GMM (Gu 和 Kanade 2008) 在第 2.2.3 和 (iv) 第三。对于 GMM, 我们经验地设置 $K_i=5$, 并使用 EM 算法来估计描述响应图的混合模型的参数。对于 RLMS, 我们使用第节中描述的有效近似。3.3.3 网格间距为 0.1 像素, 方差收紧策略为 $\rho=\{20, 10, 5, 1\}$ 。使用 (8) 中定义的线性回归器作为地标检测器。本地专家的大小为 (11×11) 像素, 在 (15×15) 像素范围内进行详尽的局部搜索窗口。混蛋, 这里比较的这些方法之间的唯一区别是响应图的近似方式以及它们的特殊优化策略。以明确这些方法: 采用 ML 公式, 假设 PDM 参数具有非信息先验。在所有情况下, 模型的比例和位置由现成的面探测器初始化, 将 (1) 中的旋转和非刚性参数设置为零 (即平均形状), 直到优化收敛为止, 测量值为: $n \quad i=1$

x_i

$2 \leq 0.01$, 其中 x_i 表示迭代之间第 i 个地标位置的变化。

这些实验的结果可以在图 5 中找到。图(拟合曲线)显示了不同水平的最大摄动的图像比例, 用实际地标和图像框中得到的 F 之间的均方根 (RMS) 误差来衡量。图例中显示了 2.5 GHz Intel Core 2 Duo 处理器上各种方法的平均安装时间。

结果表明, 四种方法的相对性能趋势一致。首先, CQF 有能力显著优于 ASM。如第 2.2.2 节所述, 这是由于 CQF 能够解释响应图中的方向不确定性以及对外围响应更为稳健。然而, CQF 有过度平滑响应图的倾向, 导致收敛精度有限。与 CQF 相比, GMM 在精度上有了提高, 收敛到的样本数量更多但是, 马勒沙佩默斯错误, 它表现出对多目标优化的敏感性。这可以从它比 CQF 更差的性能看出, 在多通道中重建误差大于 4.2 像素 RMS, 在 xm2vt 中高于 5 像素 RMS。相比之下, RLMS 比 GMM 具有更好的精度, 但仍然对局部最优保持一定程度的鲁棒性, 在局部优化中, RLMS 在严重错误初始化时的性能与 CQF 相当。最后, 尽管 RLMS 在性能上有显著的改进, 但与 ASM 和 CQF 相比, RLMS 的计算复杂度仅略有增加。这与 GMM 不同, GMM 需要更长的拟合时间, 这主要是因为将混合模型拟合到响应图的复杂性。

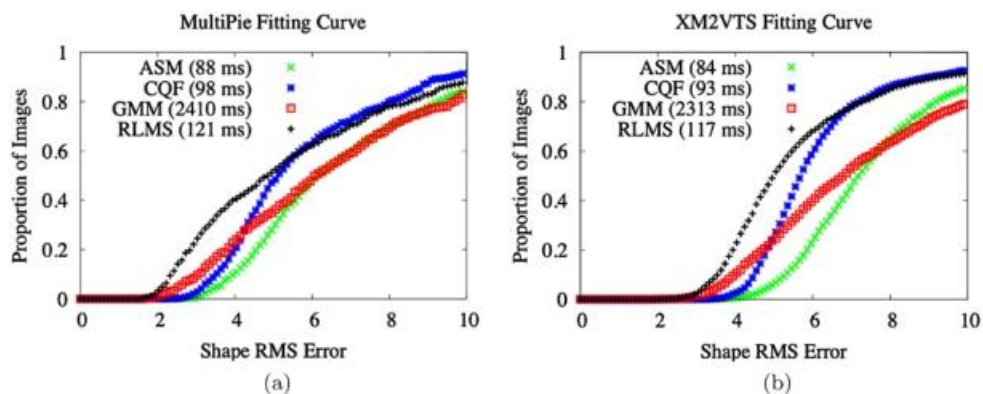


图 5 在 MultiPie 和 XM2VTS 数据库上 ASM、CQF、GMM 和 RLMS 优化策略的拟合曲线

4.1.2 序列跟踪

评估在特定数据库之外的图像上匹配算法的性能更有意义,因为它可以更好地指示方法的泛化程度。然而,这很少进行,因为它需要使用训练集的 PDM 配置来注释新图像的繁琐过程。在这里,我们利用了免费提供的 FGNet 谈话脸序列⁵。由于地面真相注释的格式与 XM2VTS 中的格式相同,因此可以对该序列进行定量分析。这里使用了静态图像实验中使用的相同模型,只是它是在 XM2VTS 中的所有图像上训练的。所有四种拟合方法均采用与前面所述相同收敛标准的 ML 公式对该序列进行评估。我们在第一帧中使用人脸检测器初始化模型,并使用前一帧中的 PDM 配置作为初始估计来初始化连续帧。虽然通过执行某种时间平滑可以提高性能,例如使用 Kalman 滤波器,但由于我们正在研究响应图估计值及其后续优化策略对拟合性能的影响,因此在这里没有执行这项工作。

在图 6 中,为四个被比较的优化策略绘制了每个帧的形状均方根误差。各种策略的相对性能与静态图像实验中观察到的相似。CQF 在整个序列中表现出良好的稳定性,但其有效性有限。GMM 在许多帧上都优于 CQF,但并不一致。此外,与 ASM 一样,它在这个序列上显得特别不稳定,在第 4200 帧左右失去轨迹,并且在序列结束之前无法恢复。GMM 拟合的不稳定本质在于它对描述响应图的混合模型的估计,这是一个依赖初始化的非线性优化过程。当我们基于 K_i 最佳响应初始化其 K_i 中心时,这可能是有噪声的,对于类似的响应映射(即来自相邻帧)可以做出非常不同的混合模型估计。类似的观察也可以用来解释 ASM 的不稳定性。相比之下,RLMS 显示出 CQF 的稳定性,但始终具有更好的精确度。

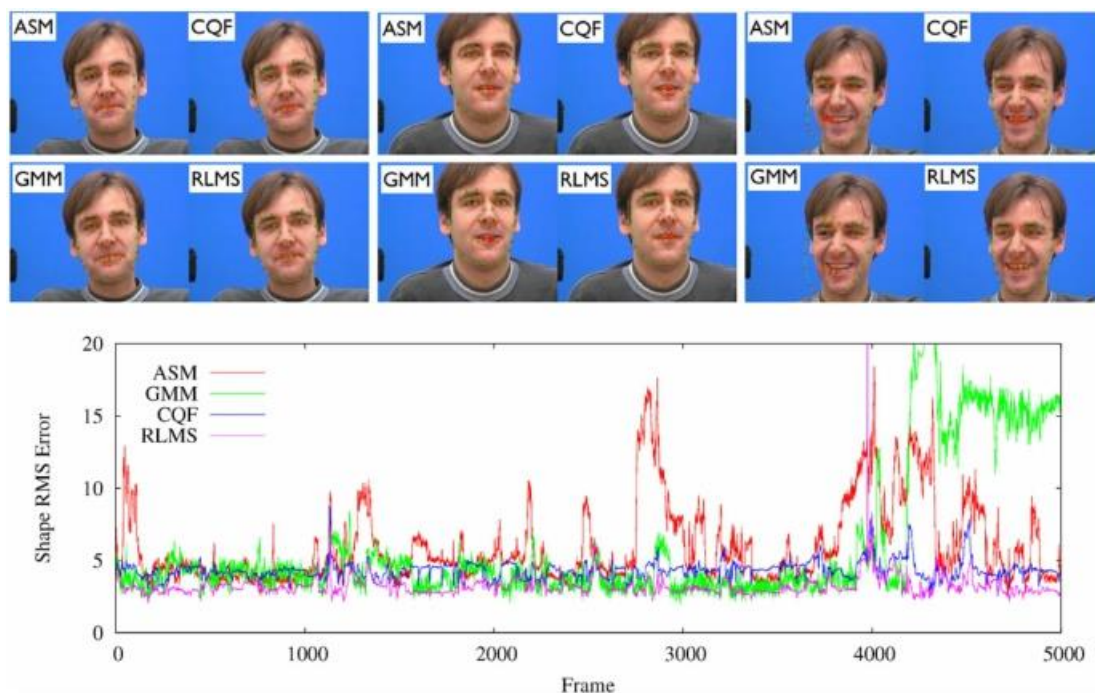


图 6 顶行：在 FGNet 谈话脸数据库上对帧 {012304200} 的跟踪结果。从左上角顺时针方向为 ASM、CQF、RLMS 和 GMM 的拟合结果。底部：整个序列中地面真相注释的形状均方根误差图

⁵http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html.

4.2 定性分析及闭塞处理

定量分析拟合算法在真实遮挡图像上的性能是困难的，因为它需要收集和注释遮挡图像。典型的评估是通过综合闭塞物 (Gross 等人, 2004 年; Saragih 2008 年)。然而，拟合方法的性能通常取决于对合成闭塞物的选择 (Saragih 2008)。此外，很难模拟遮挡对象对图像非遮挡部分的影响，例如阴影。因此，在这项工作中，我们定性地分析了遮挡对 RLMS 的影响，通过观察其在真实遮挡下的跟踪序列中的行为。

图中显示了两个完全闭塞序列的 7 帧和 8 帧，其中使用四种 RLMS 变体进行跟踪，也就是 ML 和 MAP 变量与 gaussianandgeman-McClurekernels。模特儿对整个 MultiPie 数据库进行了培训，通过将非刚性结构应用于可用注释学习 3D 形状模型 (Torresani 等人, 2008 年)。在这些序列中的闭塞类型是相当具有挑战性的，因为它们阻塞了无花果的眼睛和嘴区域。分别是 7 号和 8 号，与其他地标相比，一般会给出强烈的反应。还请注意图 7 中被遮挡物体影响的非遮挡地标上的照明变化。

在这些序列中，具有高斯核的 ML 和 MAP 变体都无法跟踪对象，当对象开始遮挡某些地标时，会立即观察到显著的扰动。使用 Geman-McClure 内核的 ML-fitting 具有更好的鲁棒性，其中遮挡的效果不是很明显。然而，误差的累积会导致在序列末尾失去跟踪。如第 3.2 节所述，仅通过地标关节运动与形状模型的一致性，使用健壮的内核强制包含不变性。因此，不考虑与形状模型一致的错误地标候选。虽然 MAP 公式采用了相同的假设，但使用先验限制了形状模型的可变性，这降低了拟合程序对这些候选对象的敏感性。使用 Geman McClure 内核进行地图匹配的跟踪结果支持这种分析，它成功地跟踪到序列的末尾。

最后，对 Wild 数据库中的人脸进行了 RLMS 的定性分析 (Huang 等人, 2007 年)。它包含在不同的光照、分辨率、图像噪声和部分遮挡下拍摄的图像。和以前一样，模型是用人脸检测器初始化的，直到收敛为止。一些拟合结果如图 9 所示。结果说明 RLMS 对真实图像中常见

的变化具有一定的鲁棒性。

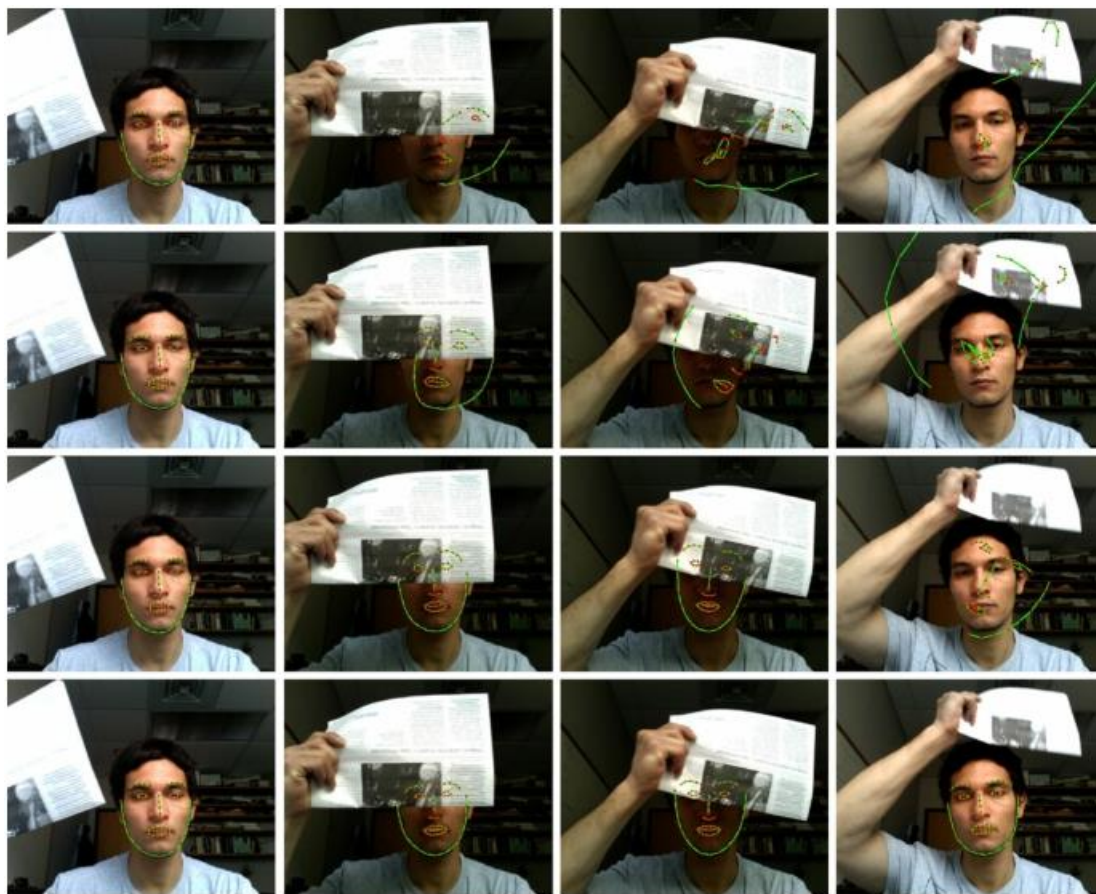


图 7 完全遮挡跟踪示例。从上到下的行：用高斯核函数拟合，用高斯核函数映射，用 Geman-McClure 核函数拟合，用 Geman-McClure 核函数映射。从左到右列：第 0、30、40 和 50 帧



图 8：完全遮挡跟踪示例。从上到下的行：用高斯核函数拟合，用高斯核函数映射，用 Geman-McClure 核函数拟合，用 Geman-McClure 核函数映射。从左到右列：第 0、30、40 和 50 帧



图 9 在 Wild 数据库中对人脸的拟合结果示例（红色矩形表示用于初始化的检测到的面部区域）。顶部：未遮挡图像。底部：部分闭塞

5 结论

本文研究了基于局部专家的可变形模型拟合的优化策略。现有的各种方法都是在一个一致的概率框架内提出的，在这个框架中，它们可以对地标位置的真实似然图进行不同的参数近似。从这个角度出发，通过数值实验解释和验证了一些常用方法的拟合行为。为了解决这些方法固有的困难，并利用从上述公式中获得的信息，提出了一种使用非参数表示的新的似然图近似。进一步的创新，降低在线计算复杂性，避免局部最优和鼓励鲁棒性的部分遮挡也被提出。结果表明，该公式的匹配算法简单、高效，并且在一般人脸匹配任务上表现出优于现有方法的性能。本文提出的方法是一个可变形模型的框架，而不是一个完整的系统。从某种意义上

说，这是一种通用的优化策略

其组件的数量可以专门用于特定的应用程序。因此，未来的工作将涉及不同特征检测器的使用（例如，见 Avidan 2004；Cootes 和 Taylor 1992；Cristinacce 和 Cootes 2006），更复杂的形状模型（参见 Gu 和 Kanade 2008；Romdhani 等人，1999 年），时间平滑约束的应用（见 Wang 等人，2008b；周等，2005），以及不同的内核类型，允许将这些内核集成到提议的框架中。

附录

计算机视觉中的许多问题很自然地需要使用图形模型来表示。这方面的例子包括立体匹配（Sun 等人，2003 年）和光学流估计（Felzenszwalb 和 Huttenlocher 2004）等。在本节中，我们将激励第 3 节中描述的 RLMS 中使用的公式。从图形模型的角度来看。无向图由一组 nodes V 和一组对应的 edges E 定义。对于分类问题，让我们定义： $V = \{y_1, \dots, y_n, p\}$ ，其中，如第 3 节所述， p 表示 PDM 参数， y_i 表示描述图像中第 i 个地标位置真实位置的随机变量。节点 $a \in V$ 的邻域定义为： $(a) = \{b \mid (a, b) \in E\}$ ，其中，对于 CLM 问题，我们有： $(y_i) = \{p\}$ 和 $(p) = \{y_i\}_{i=1}^n$ 。让对齐标签 $\{l_i\}_{i=1}^n$ 表示观察到的变量⁶，描述该系统的图采用图 10 所示的特别简单的非循环形式。由于 CLM 图是无环的，所有随机变量的条件分布可以通过一种称为信念传播（BP）的局部消息传递算法直接计算（Yedidia et al. 2002 年）。在 BP 算法的迭代 t 处，每个节点 a 计算一条消息 $m_{a \rightarrow b}$ ，发送给每个邻居 $b \in (a)$ 。对于 CLM 图形，只需发送两种类型的消息：

$$m_{y_i \rightarrow p}^t \propto \sum_{y_i \in \Psi_i} \phi(y_i, p) \psi(y_i, l_i) \quad (45)$$

$$m_{p \rightarrow y_i}^t \propto \int \phi(y_i, p) p(p) \prod_{j=1, j \neq i}^n m_{y_j \rightarrow p}^{t-1} dp, \quad (46)$$

式中 $\phi(y_i, p)$ 、 $\psi(y_i, l_i)$ 和 $p(p)$ 分别是 (31)、(6) 和 (10) 给出的各种电位。虽然 CLM 图上的 BP 算法在两次迭代中收敛，但 $m_{p \rightarrow y_i}^t$ 的计算在分析上是不可处理的（即 GMM 上的积分）。然而，由于我们对 PDM 参数的条件边缘分布很感兴趣：

⁶ 在计算机视觉问题的图形模型中，观察到的变量通常与图像像素有关。我们在这里背离这一惯例，是为了保持与第 2.1 节中对地标线形的歧视性解释保持一致，在第 2.1 节中，对图像的观察隐含在其表述中。

$$p(p \mid \{l_i = 1\}_{i=1}^n, \mathcal{I}) \propto p(p) \prod_{i=1}^n m_{y_i \rightarrow p}^t, \quad (47)$$

不需要计算消息 $\{m_{p \rightarrow y_i}\}_{i=1}^n$ ，因为它们不影响 (45) 中的消息 $\{m_{y_i \rightarrow p}\}_{i=1}^n$ 。这是假定的图结构的直接结果，其中地标检测是有条件独立的。因此，我们仅通过计算从地标候选节点到 PDM 参数节点的消息来获得期望的条件边缘分布，从而得到 (33) 中的 RLMS 目标。

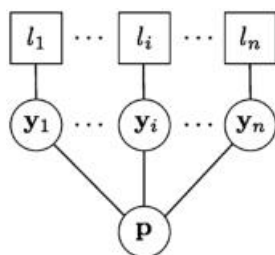


图 10: 无方向的连接图形模型。正方形表示观测值，圆圈表示隐藏变量

参考文献

- Avidan, S. (2004)。支持向量跟踪。IEEE 模式分析与机器智能汇刊 (PAMI), 261064-1072。
- Basso, C., Vetter, T. 和 Blanz, V. (2003)。正则化三维可变形模型。在 IEEE 三维建模和运动分析高级知识国际研讨会 (HLK'03) (第 3 页)。
- Black, M. 和 Anandan, P. (1993)。多个运动的稳健估计: 精细和分段平滑的流场。施乐 PARC 技术代表。
- Blake, A., Isard, M. 和 Reynard, D. (1994)。学习跟踪运动中的曲线。在 IEEE 决策理论与控制会议 (第 3788-3793 页)。
- Bruhn, A., Weickert, J. 和 Schnörr, C. (2005)。Lucas/Kanade 与 Horn/Schunck 会面: 结合局部和全局光流方法。国际计算机视觉杂志, 61 (3), 211-231。
- Carreira Perpinan, M. (2007)。高斯均值漂移是一种 EM 算法。IEEE 模式分析 (776) 和模式智能分析。
- Carreira Perpinan, M. 和 Williams, C. (2003)。关于 mod 数 Esomagussianmixture. 讲授计算机科学, 2695, 625 - 640 年。
- Cootes, T. 和 Taylor, C. (1992)。活跃的形状模型-“聪明的蛇”。在英国机器视觉会议 (BMVC'92) (第 266-275 页)。
- Cristinacce, D. 和 Cootes, T. (2004)。形状约束人脸特征检测器的比较。在 IEEE 自动人脸和手势识别国际会议 (FG'04) (第 375-380 页)。
- Cristinacce, D. 和 Cootes, T. (2006)。基于约束局部模型的特征检测与跟踪。在英国机器视觉会议 (BMVC'06) (第 929-938 页)。
- Cristinacce, D. 和 Cootes, T. (2007)。增强了活跃的形状模型。《英国机器视觉》880-880 页)。
- Dempster, A., Laird, N. 和 Rubin, D. (1977)。基于 EM 算法的不完全数据的最大似然。英国皇家统计学会期刊 B 辑 (方法论), 39 (1), 1-38。
- Edwards, G., Taylor, C. 和 Cootes, T. (1998)。使用活动外观模型解释人脸图像。在 IEEE 自动面和手势识别国际会议 (FG'98) (第 300-305 页)。

[Faceso FH1]TM

面部表情及头部运动采集模块用户手册