指导教师：杨涛　　提交时间：2016/3/17

CVPR2015 Paper Translation

No: 01

姓名：时维阳

学号：2013300220

班号：HC001310

基于二维地标的三维形状估计：凸松弛方法

Xiaowei Zhou[†], Spyridon Leonardos[†], Xiaoyan Hu[‡†], Kostas Daniilidis[†]
[†] University of Pennsylvania [‡] Beijing Normal University
{xiaowz,spyridon,kostas}@cis.upenn.edu huxy@bnu.edu.cn

摘要

我们所研究的问题是如何在给定一组单一图像的二维地标的基础上，估计出该物体的三维形状。为了缓解重建歧义，一种广泛使用的方法是把未知的三维形状建立在现有的形状空间的基础之上。虽然这种方法已被证明能够在各种应用中成功运用，但一个富有挑战性的问题仍然存在，即形状参数和摄像机位姿参数的联合估计需要解决一个非凸优化问题。解决该问题的现有方法通常采用交替最小化方案局部更新参数，因此，该解决方案对于初始化是十分敏感的。在本文中，我们提出了一个凸规划公式和一个有效的求解凸规划算法。我们验证了该方法的精确恢复性能，与其他可选的方法相比，它的优点在于该方法在人体姿态和车形状的估计方面同样具有适用性。

1. 引言

从二维图像中识别三维物体是计算机视觉中的核心问题。近年来，出现了一个新兴的趋势：分析三维几何物体的形状的和姿势而不是仅仅提供包围盒[37, 25, 4, 28, 36, 33]。三维几何推理不仅可以为后续高层次的场景提供更丰富的信息，而且还能提高目标的检测性能。[20, 31, 3, 34]。

从一个单一的视图估计物体的三维几何形状是一个具有不适定性的问题。但是这是一个观测者可能完成的任务，因为人

类可以利用物体形状的视觉记忆。在这个想法的启发下，人们已经能够开始努力利用越来越多的在线三维模型数据库进行以三维模型为基础的分析。最近的许多研究工作，采用以“主动形状模型”[14]为基础的形状空间的方法来解决类内变化或非刚性形变，例如，[22]、[32]、[45]、[35]，其中每个形状由一组有序的地标定义，同时被估计的形状被认为是预定义的一个基础形状的线性组合。为了估计形状，3D 形状模型被置于坐标系中，在图像中进行标注或者检测。这样，问题就变成了一个同时估计形状参数（线性组合的权重）和位姿参数（角度），从而完成 3D 形状由 2D 形状拟合生成的过程。

虽然这种方法在不同的应用上已取得了可喜的成果，但模型的推理仍然是一个具有挑战性的问题，由于形状和位姿估计子耦合：三维模型匹配二维地标的过程中位姿参数需要已知，然而位姿的估计需要精确的三维模型。联合估计形状和位姿参数的结果通常导致一个非凸优化问题，同时位姿参数的正交性约束将使得问题更加复杂。以前的方法通常是采用交替迭代算法，以交替更新形状和位姿参数，直到收敛。因此，该算法对初始化敏感，可能会导致局部最优解。正如许多研究中提到的，例如，[32]，[22]，大多数失败的研究实验都是由不好的初始化导致的。一些启发式算法已被用于缓解这一问题，如尝试多个初始化[35]或使用视点感知探测器来进行位姿初始化[45]。但仍不能保证全局最优。

在本文中，我们提出了凸松弛算法来解决上述问题：

1. 我们使用一个增强的形状空间模型，其中一个形状可以用一些可旋转的基础形状的线性组合来表示。该模型可以给出物体内部形状变化和外部视角变化的线性表示。

2. 我们使用正交约束的凸松弛，把整个问题转换成一个谱范数正则线性逆问题，即一个凸规划。

3. 我们开发了一个高效的算法来解决提出的全局凸规划问题。

本文的其余部分组织如下:首先，在第二节我们给出了相关研究的简介。然后，在第三节，我们就该模型做出了阐述，并第 4 节中提供相关算法。接下来，在第五节，我们通过实验证实该方法的优点和适用性。最后，在第六节中，我们用一些讨论来总结本文。

2. 相关研究

最相关的工作包括通过在 2D 地标中拟合形状空间模型的方法解决形状估计问

题。这种方法已成功地应用于各种物体的重新构造，包括人类姿态[35，18，42，24]，汽车[22，32，45，26]，面[21，12，6]，等等。以下为最近的是几个例子。

Ramakrishna 等人 [32]提出了一种基于稀疏表示的方法，在静止图像中重建三维人体姿态。Wang 等人[35]采用一个二维人体姿态探测器[40]自动定位关节，并使用一个有效的估计算法来处理不准确的关节位置。Fan 等人[18]提出在构建位姿字典时，通过加强局部特征来提高其性能。Hejrati 等人[22]采用主动形状模型进行三维汽车的重建，并通过变换部分模型的变形来产生二维地标[19]。Lin 等人[26]针对三维汽车的重建提出了一种联合 3 D 模型拟合和细粒度分类的方法。在一些研究中，将地标位置与形状拟合进行联合估计。例如，Zia [45]等人开发了一个概率框架来同时进行二维地标定位和三维物体模型恢复。Zhou 等人[42]将人的姿态估计作为一个匹配的问题 :所观测到的时空位姿模型与视频匹配从而提取轨迹。

研究该问题一个常见的部分或一个中间步骤是将 3D 模型拟合到二维地标中。在引言中提到的，以前的研究通常依赖于非凸方法，但这可能是对初始化十分敏感的。本文提出的凸规划可作为一个模块，来提高改善现有的方法的性能。

我们的研究也是与运动的非刚性结构（NRSfM）密切相关的，通过多帧 2D-2D 对应重新覆盖不断变形的形状。低阶的形状空间模型经常被用于 NRSfM，但基础形状是未知的。形状/位姿变量和基础形状的联合估计通常是通过矩阵分解后度量来矫正[9，39]。在最近的一些研究中，迭代算法有更好的精度[29，15]或顺序的处理结果[2]，因此本文研究的问题是使用类似于固定基础形状并更新剩余变量的迭代方法 NRSfM 的步骤。

3. 模型

3.1 问题阐述

本文所研究的问题可以用以下的线性系统来描述：

$$\mathbf{W} = \mathbf{\Pi S} \ , \tag{1}$$

$\mathbf{S} \in \mathbb{R}^{3\times P}$表示未知的三维形状，即由点的三维位置坐标表示。$\mathbf{W} \in \mathbb{R}^{2\times P}$表示在一个 2D 图像中的预测。Π是摄像机标定矩阵。为了简化问题，通常采用弱透视相机模型，当物体的深度远远小于距相机的距离时，可认为这一个很好的近似逼近。有了这个假设，校准矩阵具有以下简单的形式：

$$\mathbf{\Pi} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \end{pmatrix}, \tag{2}$$

α是一个取决于焦距和距物体距离的标量。

实际问题中，方程组（1）有更多的变量。为了使问题具有一定的适定性，采用一种基于主动形状模型[14]的广泛使用的假设:未知的形状可以表示为预先定义的基础形状的线性组合：

$$S = \sum_{i=1}^{k} c_i B_i, \tag{3}$$

对于$i \in [1,k], B_i \in \mathbb{R}^{3\times P}$是一个从训练数据中学习得到的基础的形状，同时$c_i$表示各基础形状的权重。通过这种方法，重建问题转化为一个估计几个拟合模型（3）在图像中的地标系数问题，从而大大降低了未知数的个数。

由于基础形状是预先定义的，所以相机架位置和定义基础形状帧之间的相对旋转和平移的需要考虑在内，3D-2D 投影可由以下公式表示：

$$\mathrm{W} = \Pi_{\mathrm{S}}\left(\mathrm{R}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{c_i B_i} + \mathrm{T1^T}\right), \tag{4}$$

$\mathbf{R} \in \mathbb{R}^{3\times3}$和$\mathbf{T} \in \mathbb{R}^{3}$分别对应旋转矩阵和平移向量。R 应在一个特殊的的正交群中

$$\mathrm{SO(3)} = \{\mathrm{R} \in \mathbb{R}^{3\times3} | \mathrm{R^T R} = \mathrm{I_3}, \det \mathrm{R} = 1\}. \tag{5}$$

方程（4）可以进一步简化为

$$\mathbf{W} = \bar{\boldsymbol{R}} \sum_{i=1}^{k} c_i B_i, \tag{6}$$

其中$\bar{\boldsymbol{R}} \in \mathbb{R}^{2\times3}$代表旋转矩阵的前两行，同时平移量$\boldsymbol{T}$可以通过集中数据的方法消除，即用每一行的平均值减去$\boldsymbol{W}$和$\boldsymbol{B}$。需要注意的是，在校准矩阵中的标量α已被并入到$c_i, \ldots, c_k$中。

在主动形状模型中，基础形状的数目被设定的相对较小，也就是假定未知的形状是在一个低维线性空间中。最近的许多研究[32，41，43，44]，表明低维线性空间不能满足拟合复杂模型的形状变化的需求，例如，人的姿势。一个有研究价值的方法是使用一个完整的字典，表示一个未知的形状为一个字典中元素的稀疏组合。这种稀疏表示隐式编码的假设，使得未知的形状近似表示为一组非线性形状的多样组合的子空间。

基于采用稀疏表示表示形状，后续的优化问题往往被认为是一个未知形状的估计问题：

$$\begin{aligned} &\min_{\boldsymbol{c},\bar{\boldsymbol{R}}} \ \frac{1}{2}\left\| \boldsymbol{W} - \bar{\boldsymbol{R}} \sum_{i=1}^{k} c_i \boldsymbol{B}_i \right\|_F^2 + \lambda \|\boldsymbol{c}\|_1, \\ &\text{s.t.} \ \ \bar{\boldsymbol{R}}\bar{\boldsymbol{R}}^T = \boldsymbol{I}_2, \end{aligned} \tag{7}$$

其中$\boldsymbol{c} = [c_i, \ldots, c_k]^T$和$\|\boldsymbol{c}\|_1$表示$\boldsymbol{c}$的范数$\boldsymbol{\ell_1}$，即基数的凸替代。$\|\cdot\|_F$表示一个矩阵的 Frobenius 范数。公式（7）分别表示代价函数对应的重投影误差和稀疏形式。

公式（7）中的优化是非凸的，有一个正交性约束。一个常用的策略是交替迭代最小化算法，其中两步运用交替迭代：通

过求解ℓ_1的最小化问题固定$\bar{R}$并更新c的值；固定c并使用一定的旋转表示如四元数，指数映射或流形表示来更新$\bar{R}$。需要注意的是 Procrustes 方法不能直接应用在这里，因为$\bar{R} \in \mathbb{R}^{2\times3}$不是一个完整的旋转矩阵，一般不存在封闭的解决方案[17]。因此，整个算法可能会陷入局部最小，从而远离全局最优解。

3.2 提出模型

我们提出了下面的形状空间模型：

$$S = \sum_{i=1}^{k} c_i R_i B_i, \tag{8}$$

每一个基础形状都有一个旋转量。公式（8）隐式地解释了视角变化，投影的二维模型表示为：

$$W = \Pi \sum_{i=1}^{k} c_i R_i B_i = \sum_{i=1}^{k} M_i B_i, \tag{9}$$

$M_i \in \mathbb{R}^{2\times3}$产生于$c_i$和$R_i$的前两行，满足

$$M_i M_i^T = c_i^2 I_2. \tag{10}$$

使用模型（8）和（9）的目的是实现线性表示二维的形状变化，这样，我们可以摆脱双线性形式（6）。对于凸规划来讲这是一个必要的步骤。

模型（9）与现有的研究[38，5]中的仿射形状模型等价，它使用一个增强的线性空间来表示二维的形状变化引起的物体内部的形状变形和外部的视点变化。这种表示也出现在大多数有关 NRSfM 的研究中[9，29]。正如在[38]中所提到的，增强的线性空间可以表示由三维形状模型投影到图像平面产生的任何二维形状，但自由度的增加，可能会导致无效的形状。在这项研究工作中，我们尝试通过加强在基形状的数目上执行的正交约束和稀疏约束，以减少产生无效的形状的可能性。我们将证明，这些限制可以方便的通过最小的凸化方法施加。

接下来，在（10）中我们将通过其凸对应考虑正交约束。以下引理已被证明在文献[23，3.4]：

引理 1. Stiefel 流型凸包 $Q = \{X \in \mathbb{R}^{m\times n} | X^T X = I_n\}$等于单位谱范数球转换 $(Q) = \{X \in \mathbb{R}^{m\times n} | \| X \|_2 \le 1\}$. $\| X \|_2$指矩阵 X 的谱范数（又名诱导范数），它被定义为 X 的最大奇异值。

基于引理 1，我们有以下命题：

命题 1. 给定一个标量 s，它的凸包 $s = \{Y \in \mathbb{R}^{m\times n} | Y^T Y = s^2 I_n\}$等于半径为 $|s|$ 的谱范数球。

$$\text{conv}(S) = \{Y \in \mathbb{R}^{m\times n} | \| Y \|_2 \leqslant |s|\}$$

$$M_i M_i^T = c_i^2 I_2. \tag{10}$$

证明是简单的，因为S和Q之间有一个$Y = sX$的线性映射关系。

因此，通过$\|M_i\|_2 \le |C|_i$在（10）中给出了严格的凸松弛的约束。

最后，我们使用改进的松弛正交性约束和假定稀疏表示的形状模型，采用减少在无噪声的情况下恢复系数向量的 ℓ1 范数

$$\min_{c_1,\cdots,c_k,M_1,\cdots,M_k} \sum_{i=1}^{k} |c_i|,$$
$$\text{s.t.} \quad W = \sum_{i=1}^{k} M_i B_i,$$
$$\|M_i\|_2 \le |c_i|,\ \forall i \in [1,k] \qquad (11)$$

这显然等价于以下问题：

$$\min_{M_1,\cdots,M_k} \sum_{i=1}^{k} \|M_i\|_2,$$
$$\text{s.t.} \quad W = \sum_{i=1}^{k} M_i B_i. \qquad (12)$$

（12）中的规划是通过最小化谱范数来估计一组正交矩阵的逆线性问题，。有趣的是，使用该凸规划的精确恢复条件已在[13]中进行了理论上的证明。我们将提供计算结果，以证明在第 5.1 节的准确恢复特性。

考虑实际应用中的噪声，我们可以解决：

$$\min_{M_1,\cdots,M_k} \frac{1}{2} \left\| W - \sum_{i=1}^{k} M_i B_i \right\|_F^2 + \lambda \sum_{i=1}^{k} \|M_i\|_2. \qquad (13)$$

我们最后规划形式（13）是一个惩罚最小二乘问题。我们有以下备注：

1. (13) 是可以在全局范围内得到解决的凸规划问题。我们将在第四节提供一个有效的算法。

2. 注意$\|\cdot\|_2$在上述公式表示一个矩阵的谱范数而不是一个向量范数 ℓ2。正如我们将在 4 节中展示的，一个矩阵的谱范数极小化等价于最小化的奇异值向量的范数ℓ_∞，这将同时收缩矩阵的范数向零，并使其奇异值相等。因此，通过谱范数的最小化，我们不仅可以减少使用的基础形状的数目，也可以使每个变换矩阵是正交的（正交矩阵具有相等的奇异值）。

3. 在实践中，我们通过考虑可见地标的重投影误差可以估计$M_i s$，即（13）的第一部分的一个二进制权重矩阵。因为其位置是已知的基础形状，所以丢失的地标可以在重建的形状模型的基础上得到。

3.3. 重建

(13) 解决后，我们通过估计M_i恢复C_i和R_i，并通过 (8) 重建其三维形状。具体而言就是$c_i = \|M_i\|_2$和$\bar{R}_i = M_i / c_i$。注意，$c_i = -\|M_i\|_2$也是一个可行的解决方案。为了消除歧义，我们假设认为$c_i \ge 0$，在训练形状字典时施加该约束。最后，R_i的第三行是由$\bar{R}_i$中行的交叉乘积产生。

4. 优化

4.1. 谱范数逼近算子

在使用特定的算法来解决（13）之前，我们首先证明以下命题，这将作为我们的算法的一个重要组成部分。

命题 2. 以下问题的解决方案

$$\min_X \frac{1}{2} \|Y - X\|_F^2 + \lambda \|X\|_2 \qquad (14)$$

$$\mathcal{D}_\lambda(Y) = U_Y\, diag\,[\sigma_Y - \lambda \mathcal{P}_{\ell_1}(\sigma_Y/\lambda)]\, V_Y^T, \qquad (15)$$

由式子 $X^* = \mathcal{D}_\lambda(Y)$, 给出 U_Y, V_Y和 σ_Y分别表示Y的左奇异向量、右奇异向量和奇异值。$\mathcal{P}_{\ell_1}$是矢量到单位 ℓ 范数球的投影。

证明. 问题（14）是一个近端问题[30]。与函数F相关联的近端问题定义如下：

$$\text{prox}_{\lambda F}(Y) = \arg\min_X \frac{1}{2}\|Y - X\|_F^2 + \lambda F(X), \quad (16)$$

用$\mathcal{P}_{\lambda F}(Y)$表示该解决方案并定义邻近算子$F$.

针对问题(14),$F(X) = \|X\|_2 = \|\sigma_X\|_\infty$ ，其中$\|\cdot\|_\infty$表示ℓ_∞的范数。F是一个作用于矩阵的奇异值的谱函数。基于谱函数的性质[30，第 6.7.2 节]，可得

$$\text{prox}_{\lambda F}(Y) = U_Y \,\text{diag}\left[\text{prox}_{\lambda f}(\sigma_Y)\right] V_Y^T, \quad (17)$$

其中，f表示ℓ_∞-的范数。ℓ_∞-范数的临近算子可以由 Moreau 进行分解[30，第 6.5 节]：

$$\text{prox}_{\lambda f}(\sigma_Y) = \sigma_Y - \lambda \mathcal{P}_{\ell_1}(\sigma_Y/\lambda), \quad (18)$$

可以得出ℓ_1范数是ℓ_∞-范数的对偶范数。

4.2 算法

我们提出算法来解决（13）。在无噪声的情况下，（12）也可以用同样的方法解决。我们的算法实现基于交替方向乘子法（ADMM）[8]和近似算子的谱范数。

我们首先引入一个辅助变量 Z 并重写（13）如下

$$\min_{\widetilde{M},Z} \frac{1}{2}\left\|W - Z\widetilde{B}\right\|_F^2 + \lambda\sum_{i=1}^k \|M_i\|_2,$$
$$\text{s.t. } \widetilde{M} = Z, \quad (19)$$

我们定义$\mathbf{M}_1, \cdots, \mathbf{M}_K$为$\widetilde{M}$的列向量，$\mathbf{B}_1, \cdots, \mathbf{B}_K$为$\widetilde{B}$的行向量

（19）的增强拉格朗日表示形式为：

$$\mathcal{L}_\mu\left(\widetilde{M}, Z, Y\right) = \frac{1}{2}\left\|W - Z\widetilde{B}\right\|_F^2 + \lambda\sum_{i=1}^k \|M_i\|_2 + \left\langle Y, \widetilde{M} - Z\right\rangle + \frac{\mu}{2}\left\|\widetilde{M} - Z\right\|_F^2, \quad (20)$$

其中 Y 是一个双变量，μ是一个控制优化步长的参数。然后进行，交替的拆解直到收敛：

$$\widetilde{M}^{t+1} = \arg\min_{\widetilde{M}} \mathcal{L}_\mu\left(\widetilde{M}, Z^t, Y^t\right); \quad (21)$$
$$Z^{t+1} = \arg\min_{Z} \mathcal{L}_\mu\left(\widetilde{M}^{t+1}, Z, Y^t\right); \quad (22)$$
$$Y^{t+1} = Y^k + \mu\left(\widetilde{M}^{t+1} - Z^{t+1}\right). \quad (23)$$

对于(21)中的步骤，我们可得：

$$\begin{aligned}&\min_{\widetilde{M}} \mathcal{L}_\mu\left(\widetilde{M}, Z^t, Y^t\right)\\ &= \min_{\widetilde{M}} \frac{1}{2}\left\|\widetilde{M} - Z^t + \frac{1}{\mu}Y^t\right\|_F^2 + \frac{\lambda}{\mu}\sum_{i=1}^k \|M_i\|_2\\ &= \min_{M_1,\cdots,M_k} \sum_{i=1}^k \left\{\frac{1}{2}\|M_i - Q_i^t\|_F^2 + \frac{\lambda}{\mu}\|M_i\|_2\right\}, \quad (24)\end{aligned}$$

其中Q_i^t是$Z^t - \frac{1}{\mu}Y^t$的第i列，因此，我们可以通过解决基于命题 2 的近端问题来更新M_i:

$$M_i^{t+1} = \mathcal{D}_{\frac{\lambda}{\mu}}(Q_i^t), \ \forall i \in [1, k]. \quad (25)$$

对于(22)中的步骤，$\mathcal{L}_\mu(\tilde{M}^{t+1}, Z, Y^t)$是一 Z 的个二次形式，并具有以下封闭形式解：

$$Z^{t+1} = \left(W\tilde{B}^T + \mu\tilde{M}^{t+1} + Y^t\right)\left(\tilde{B}\tilde{B}^T + \mu I\right)^{-1}. \quad (26)$$

它可以证明在（21）到（23）中由 ADMM 迭代产生的值序列，收敛至原问题（19）[8]的最优解，在（19）[8]，这也是原问题（13）的最优解。

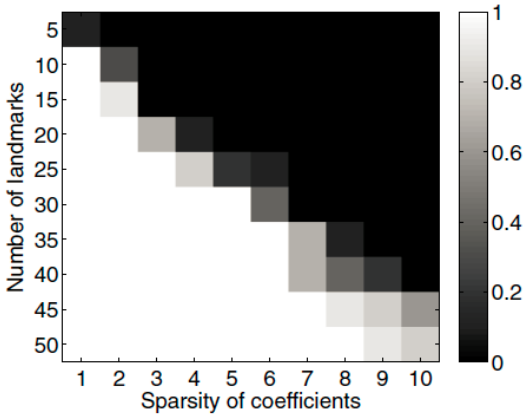

图 1. 合成数据的精确恢复频率

5. 实验

5.1. 模拟仿真

我们的目的是探讨谱范数最小化（12）是否可以解决基于无噪声的情况下的稀疏性和正交性的先验知识的不适定逆问题。

更具体地说，我们随机模拟 K 个输入采样独立并遵循（0，1）正态分布基础形状$\mathbf{B}_1, \cdots, \mathbf{B}_K \in \mathbb{R}^{3\times p}$（p 是变化的，K= 50），同时模拟旋转矩阵$\mathbf{R}_1, \cdots, \mathbf{R}_K$以及系数$\mathbf{C}_1, \cdots, \mathbf{C}_K$。只有随机选择的系数 Z 是符合（0，1）均匀采样分布的非零的值。接下来，得$\mathbf{M}_i = \mathbf{c}_i\bar{R}_i \in \mathbb{R}^{2\times 3}$和$W = \sum_{i=1}^{k} M_i B_i$，我们使$W$作为输入，解决(12)来估计$M_i s$。如果$\|\widehat{M} - \tilde{M}\|_F / \|\tilde{M}\|_F < 10^{-3}$，其中，连接$\mathbf{M}_i s$形成$\tilde{M}$，$\widehat{M}$是算法估计量。

图 1 为在变量p（地标的数量）和z（稀疏的基本系数）两个变量作用下的精确恢复的频率，这是通过评估超过 10 个随机生成的实例得出的。请注意，未知数的数目（6K）是远远大于方程的数目（2P）。本文提到的凸规划可以完全解决该下三角形区域频率等于 1 的问题，在下三角区域在地标数目足够大的并且系数是稀疏的。这表明凸松弛的作用，它被证明在各种逆问题的处理中是可行的，例如，压缩感知[11]和完成矩阵[10]。在更复杂的情况下上三角区的性能将会下降。这一现象类似于压缩感知的相变，其恢复概率也取决于观测结果和底层的信号稀疏度[16]。

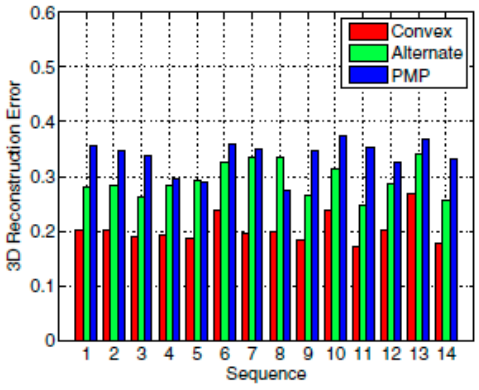

图 2. 15 个 d 动作的动作捕捉数据集序列的平均重构误差。用三种方法进行了比较：“凸”表示凸规划的方法；“交替”方式表示交替最小化方法；“PMP”代表[32]中提到的方法。

	Convex	Alternate	PMP
Subject 13	0.259	0.293	0.390
Subject 14	0.258	0.308	0.393
Subject 15	0.204	0.286	0.340

表 1. 三名个体的动作捕捉数据集序列的平均误差

5.2. 应用

5.2.1人体姿态估计

在以往的工作中，稀疏的形状表示的三维人体姿势恢复的适用性已经进行了深入的研究[32，35，18]。在本文中，我们的目标是说明提出的的凸规划与以前的研究中广泛使用的交替最小化的相比的优点。

我们对动作捕捉数据集[1] 进行评价，使用主体86的序列作为训练数据，使用从主题13、14、15提取的序列作为测试数据。所有选定的个体进行了大量的运动，如跑步，跳跃，拳击，篮球等，因为有成千上万的训练形状，使用所有的基本形状是不切实际的。对于我们的方法而言，我们解决以下问题来建立一个学习形状字典：

$$\min_{B_1,\cdots,B_k,C} \sum_{j=1}^{n} \frac{1}{2}\|S_j - \sum_{i=1}^{k} C_{ij}B_i\|_F^2 + \beta\sum_{i,j} C_{ij}$$
$$\text{s.t. } C_{ij} \geq 0,\ \|B_i\|_F \leq 1,$$
$$\forall i \in [1,k],\ j \in [1,n], \qquad (27)$$

$\mathbf{B}_i s$是学习的基础形状，$\mathbf{S}_i s$表示训练形状的（Procrustes方法对齐），$\mathbf{C}_{ij}$表示第j个训练形状的第i个系数。

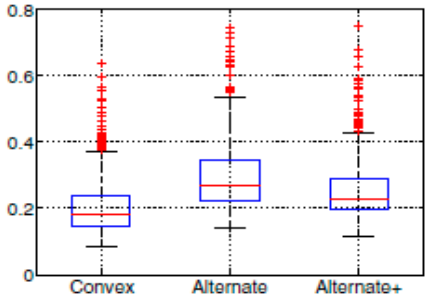

图 3. 估计误差对运动捕捉数据的盒装图（主体 15）的方法（“凸”），交替最小化（“候补”）和交替最小化由凸方法初始化（“交替+”）

我们一种广泛使用在字典学习中的策略[27]：均匀地从训练数据选择K种形状，通过交替更新$\mathbf{C}$和$\mathbf{B}_i s$的本地解决方案（27）来初始化字典。我们使用如图4所示的15个关节模型，并设置K为64

我们将我们所提出的方法与Ramakrishna等人提出的预计匹配追踪（PMP）[32][1]相对比。我们还实现了一个交替最小化方法，通过使ℓ1最小化和流形优化来更新位姿参数$\bar{R}$来交替更新的形状参数$\boldsymbol{c}$，从而解决模型（7）中的问题。该流形分析是由Stiefel流形的信赖域算流形，使用manopt工具箱[7]来更新$\bar{R}$。交替最小化是通过训练形状的平均形状初始化的。对于所提出的方法和交替最小化方法，我们为所有序列设置正则化参数为λ = 0.1。

重建误差是通过重建的形状和真实形状的相似变换的欧氏距离来进行评估的。14个测试序列在15个测试中的平均误差如图2所示。这个主题是在不同的序列中进行各

1该代码可以通过作者的网站下载：
http://www.cs.cmu.edu/~vramakri/research.html

种各样的活动[1]。凸算法显然优于替代方法，并对所有序列都具有稳定性。表1中给出了每个对象的所有序列的平均误差。

要验证的交替最小化取决于初始化，我们用我们的方法的解决方案初始化交替最小化。主题15的结果如图3所示。通过使用更好的初始化，有较小的方差，交替最小化的误差明显下降。交替最小化和无凸初始化平均目标值分别为0.17和0.24[2]。“替代+”的准确性比凸规划差。这可能使得在（8）的形状模型比原来模型（3）提供更多的自由度来表示复杂的变形的人类骨骼。

三个选定的帧的重建提出了可视化的图4。我们可以看到，所有的方法都表现良好，在第一个例子中，形状（步行）是接近的平均形状（直）。但替代方法的精度降低在其他2个例子，其中的形状是远离的平均形状，而我们的方法仍然得到有吸引力的重建。

三个选定的帧的重建姿态如图4。我们可以看到，所有的方法都表现良好，在第一个例子中，形状（步行）接近的平均形状（直立）。但替代方法的精度降低在其他2个方法，其形状远离平均形状，而我们的方法能得到有较好的重建效果。

2凸规划的目标是不同的，因此不比较。

5.2.2汽车重构

我们使用最近公布的精确的三维汽车数据集[26]证明了所提出的方法的适用性，数据集中提供了汽车的图像、二维地标、标注和相应的三维模型。我们将15辆车的3D模型作为形状字典，试图从可见的地标图像的标注中重建其他汽车的三维模型（每个图像有40个标注点）。在数据集中作者提供的图像为三维模型重建图像，而不是真正的计算机辅助设计模型。因此，我们只展示了一些定性的结果。如图5所示，我们的方法可以成功地构建各种模型如轿车、SUV和皮卡。为了进行比较，我们在原来的图像上显示了替代方法的结果[26]，它使用了透视相机模型和非线性优化的方法。替代方法是利用平均形状初始化，在轿车的例子效果良好，但在模型偏离平均形状的SUV和卡车的例子中相对较差。类似的结果也在先前的论文[26]中提及，作者提出使用特定
类的平均形状来进行更好的初始化。相反，我们的方法可以实现在任意初始化条件下结果都十分令人满意的效果。

5.3. 计算时间

我们的算法在Matlab中实现和测试，计算机配置为英特尔i7 3.4GHz处理器，8G内存。在我们的实验中，ADMM算法一般在500个迭代内收敛，大约10-4。 在人体姿态估计的实验中，我们的算法的计算时间是每帧0.33s，而交替最小化和PMP算法[32] 分别是0.44s和3.02s。

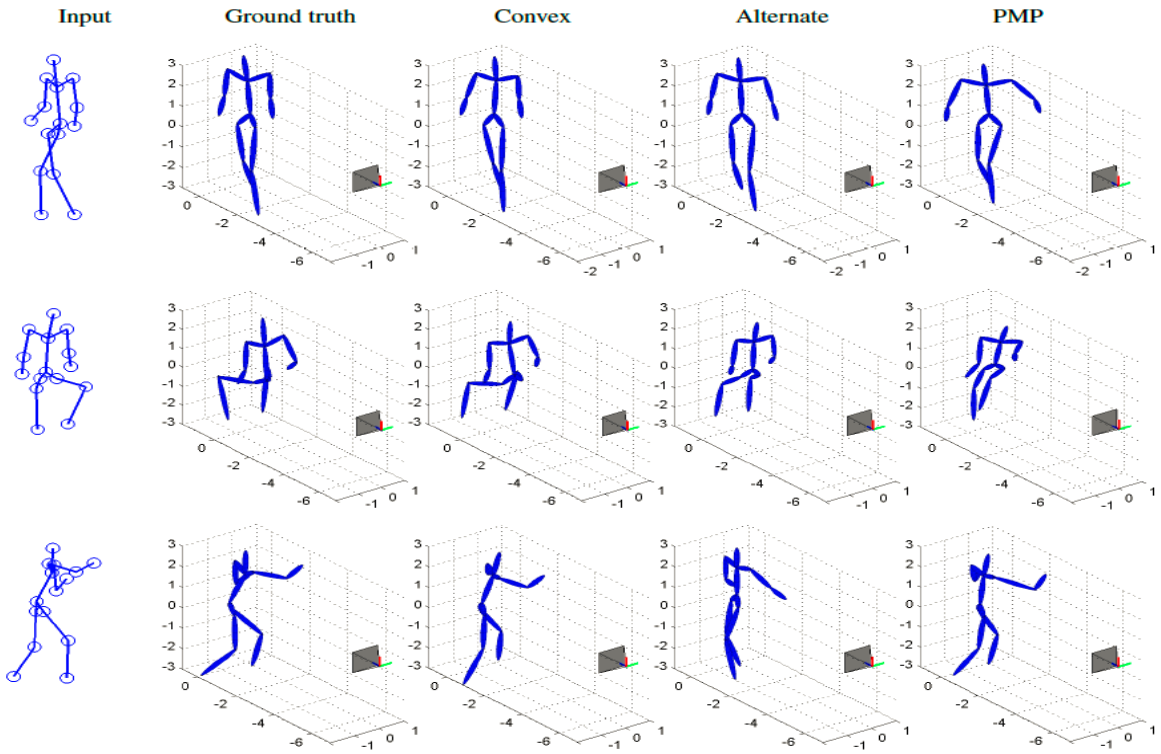

图4. 人体姿态估计实例。从左到右的列分别对应于输入的2D地标，真实的三维姿态，重建的方法产生的三维姿态，交替最小化产生的三维姿态，和PMP方法[32] 产生的三维姿态

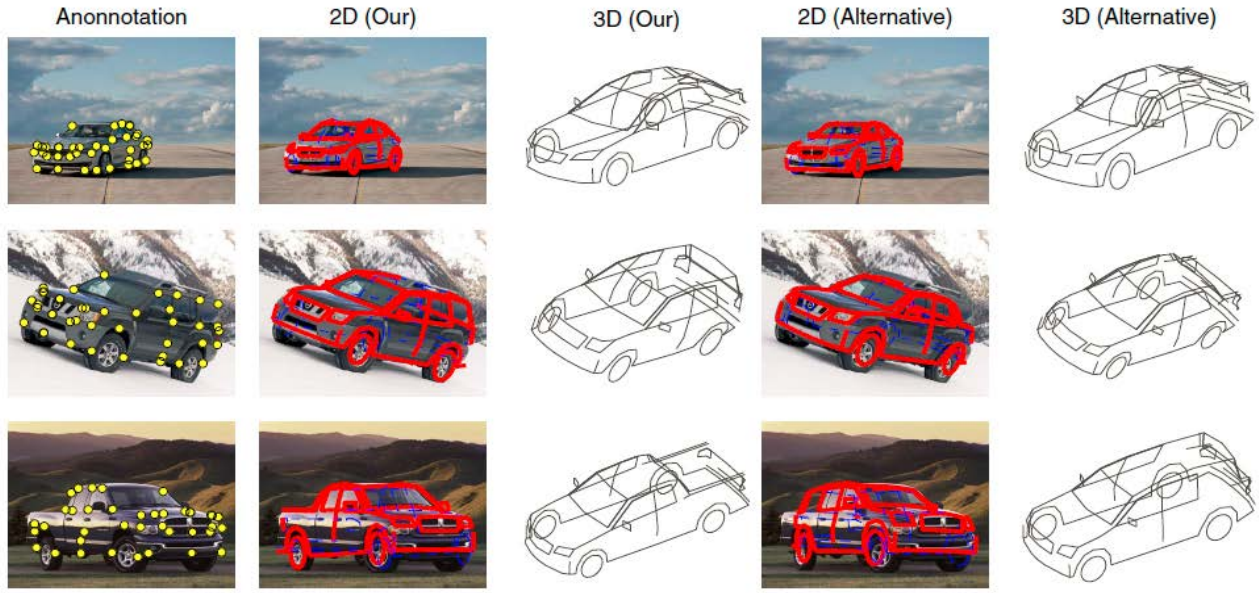

图5.汽车重构实例。从左到右各列对应于输入的二维地标，二维拟合模型和所提出的方法的三维重建、替代方法的结果[26]，分别。唯一可见的地标（~40/每幅图像）用于形状拟合。三维模型是可视化。从上到下的汽车模型分别是宝马5系列2011（轿车），日产Xterra 2005（SUV）和道奇Ram2003（卡车）。

6. 讨论

总之，我们提出了一种方法，通过求解一个凸规划将一个三维形状空间模型调整为相应的二维地标，并保证全局最优。通俗地讲，我们采用增强的 3D 形状模型来实现二维形状变化的线性变化，同时提出使用谱范数正则化来避免由谱范数增大引起的无效情况。

在各种假设下，例如逆问题的正确性、稀疏性和正交性等，利用已有文献例如[13]，从理论上分析了利用线性凸松弛解决问题的正确性。在实验中，我们观察到，在大多数情况下，估计在大部分情况下满足原来的约束，但所有研究显示该算法的输出没有细化。在松弛不紧的情况下，后续处理步骤可以用来提高正确性，例如，将估算的旋转矩阵至（3）或迫使基础形状旋转相同的角度。虽然我们没有在实验中使用这种方法，但在实际应用中建模刚性物体时，这可能是有用的。

在本文中，我们假设 2D 地标和 3D-2D 的对应关系已经给出。我们的方法可以很自然地处理较大的错误，这在实践中具有里程碑意义。例如，ℓ 范数可以用来代替（13）中的平方损失，使模型对离群点的鲁棒性提高，优化也可以使用 ADMM 算法。另一个可能的解决方案是使用[24]提出的 RANSAC，因为可以通过只使用一部分地标来估计形状模型。此外，将所得的形状模型与现有的地标定位方法相结合，同时定位二维地标、恢复形状的方法具有一定的研究价值。

致谢：感谢以下补助支持：NSF-DGE-0966142,NSF-IIS-1317788,NSF-IIP-1439681,NSF-IIS-1426840,ARLMAST-CTAW911NF-08-2-0004,ARLRCTAW911NF-10-2-0016,ONRN000141310778.XiaoyanHu获国家自然科学基金项目（no.61103086和61170186）支持。

参考文献

[1] Mocap: Carnegie mellon university motion capture database. http://mocap.cs.cmu.edu/. 6

[2] A. Agudo, L. Agapito, B. Calvo, and J. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In CVPR, 2014. 2

[3] M. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estima- tion and tracking by detection. In CVPR, 2010. 1

[4] M. Aubry, D. Maturana, A. Efros, B. Russell, and J. Sivic. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In CVPR, 2014. 1

[5]A. Blake and M. Isard. Active contours. Springer, 2000. 3

[6] V. Blanz and T. Vetter. Face recognition based on fitting a 3D mor- phable model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(9):1063–1074, 2003. 2

[7] N.Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a matlab toolbox for optimization on manifolds. Journal of Machine Learning Research, 15:1455–1459, 2014. 6

[8] S. Boyd. Distributed optimization and statistical learning via the al- ternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1):1–122, 2010. 5

[9] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In CVPR, 2000. 2, 3

[10] E. J. Cande`s and T. Tao. The power of convex relaxation: Near- optimal matrix completion. IEEE Transactions on Information The- ory, 56(5):2053–2080, 2010. 5

[11] E. J. Cande`s and M. B. Wakin. An introduction to compressive sam- pling. IEEE Signal Processing Magazine, 25(2):21–30, 2008. 5

[12] C. Cao, Y. Weng, S. Lin, and K. Zhou. 3d shape regression for real-time facial animation. ACM Transactions on Graphics (TOG), 32(4):41, 2013. 2

[13] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. Foundations of Compu- tational Mathematics, 12(6):805–849, 2012. 4, 8

[14] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models

– their training and application. Computer Vision and Image Under- standing, 61(1):38–59, 1995. 1, 2

[15] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini. Bilinear mod- eling via augmented lagrange multipliers (balm). IEEE Transac- tions on Pattern Analysis and Machine Intelligence, 34(8):1496– 1508, 2012. 2

[16] D. Donoho and J. Tanner. Observed universality of phase transi- tions in high-dimensional geometry, with implications for modern data analysis and signal processing. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sci- ences, 367(1906):4273–4293, 2009. 5

[17] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algo- rithms with orthogonality constraints. SIAM journal on Matrix Anal- ysis and Applications, 20(2):303–353, 1998. 3

[18] X. Fan, K. Zheng, Y. Zhou, and S. Wang. Pose locality constrained representation for 3d human pose reconstruction. In ECCV, 2014. 2, 6

[19] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In CVPR, 2008. 2

[20] S. Fidler, S. Dickinson, and R. Urtasun. 3d object detection and view- point estimation with a deformable 3d cuboid model. In Advances in Neural Information Processing Systems, 2012. 1

[21] L. Gu and T. Kanade. 3D alignment of face in a single image. In CVPR, 2006. 2

[22] M. Hejrati and D. Ramanan. Analyzing 3d objects in cluttered im- ages. In Advances in Neural Information Processing Systems, 2012. 1, 2

[23] M. Journe´e, Y. Nesterov, P. Richta´rik, and R. Sepulchre. Generalized power method for sparse principal component analysis. The Journal of Machine Learning Research, 11:517–553, 2010. 3

[24] Y. Li, L. Gu, and T. Kanade. Robustly aligning a shape model and its application to car alignment of unknown pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(9):1860–1876, 2011. 2, 8

[25] J. J. Lim, H. Pirsiavash, and A. Torralba. Parsing ikea objects: Fine pose estimation. In ICCV, 2013. 1

[26] Y.-L. Lin, V. I. Morariu, W. Hsu, and L. S. Davis. Jointly optimizing 3d model fitting and fine-grained classification. In ECCV, 2014. 2, 7, 8

[27] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and

sparse coding. The Journal of Machine Learning Research, 11:19–60, 2010. 6

[28] G. Mori and J. Malik. Recovering 3d human body configurations using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(7):1052–1062, 2006. 1

[29] M. Paladini, A. Del Bue, J. Xavier, L. Agapito, M. Stos˘ic´,and M. Dodig. Optimal metric projections for deformable and articu-lated structure-from-motion. International Journal of Computer Vi- sion, 96(2):252–276, 2012. 2, 3

[30] N. Parikh and S. Boyd. Proximal algorithms. Foundations and Trends in Optimization, 1(3):123–231, 2013. 4

[31] B. Pepik, P. Gehler, M. Stark, and B. Schiele. 3d2pm–3d deformable part models. In ECCV, 2012. 1

[32] V. Ramakrishna, T. Kanade, and Y. Sheikh. Reconstructing 3d human pose from 2d image landmarks. In ECCV, 2012. 1, 2, 3, 6, 7, 8

[33] L. Sigal, M. Isard, H. Haussecker, and M. J. Black. Loose-limbed people: Estimating 3d human pose and motion using non-parametric belief propagation. International Journal of Computer Vision, 98(1):15–48, 2012. 1

[34] E. Simo-Serra, A. Quattoni, C. Torras, and F. Moreno-Noguer. A joint model for 2d and 3d pose estimation from a single image. In CVPR, 2013. 1

[35] C. Wang, Y. Wang, Z. Lin, A. L. Yuille, and W. Gao. Robust estima- tion of 3d human poses from a single image. In CVPR, 2014. 1, 2, 6

[36] X. K. Wei and J. Chai. Modeling 3d human poses from uncalibrated monocular images. In ICCV, 2009. 1

[37] Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012. 1

[38] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+ 3d active appearance models. In CVPR, 2004. 3

[39] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non- rigid shape and motion recovery. International Journal of Computer Vision, 67(2):233–246, 2006. 2

[40] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, 2011. 2

[41] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Sparse shape composition: A new framework for shape prior model- ing. In CVPR, 2011. 3

[42] F. Zhou and F. De la Torre. Spatio-temporal matching for human detection in video. In ECCV, 2014. 2

[43] S. Zhu, L. Zhang, and B. M. Smith. Model evolution: an incremental approach to non-rigid structure from motion. In CVPR, 2010. 3

[44] Y. Zhu, D. Huang, F. De la Torre Frade, and S. Lucey. Complex non-rigid motion 3d reconstruction by union of subspaces. In CVPR, 2014. 3

[45] M. Z. Zia, M. Stark, B. Schiele, and K. Schindler. Detailed 3d repre- sentations for object recognition and modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence,35(11):2608–2623, 2013. 1, 2