

信号与数据处理中的低秩模型——理论、算法与应用*

林宙辰¹ 马毅²

¹ 北京大学机器感知与智能教育部重点实验室, 北京 100871

² 上海科技大学信息学院, 上海 201210

1. 引言

我们正处于大数据时代。大数据的特点,除了大家津津乐道的4个V (Volume, Variety, Velocity, 和 Value) 外也经常是高维的。例如,用千万像素相机拍一张照片,就是在千万维空间里采了一个样本。因此,如何鲁棒高效地处理高维数据是一个重大的挑战。幸运的是,高维数据并不是毫无结构的。一个显著的特点是它们经常分布在低维流形附近,这事实上就是流形学习的基本假设。这个特点也经常被事实所验证,比如对数据做主成分分析 (Principal Component Analysis, PCA) 时,往往95%以上的能量都集中在少数几个主方向上,这就说明数据是分布在低维子空间附近的。因此,充分利用数据分布的低维特性来鲁棒高效地处理高维数据是一个重要的途径。传统的流形学习属于这个范畴,但是传统的流形学习对方法的正确性缺乏深入的理论分析,因此其性能没有保障,尤其是当数据带强噪声、有损毁或缺失时甚至可能失效,而这样的数据是现实中大量存在的。

基于低秩的模型是近年涌现出来的鲁棒高效地处理高维数据的新工具。虽然秩在统计学中早已被用作矩阵的正则化子,如减秩回归 (Reduced Rank Regression, RRR); 在三维立体视觉里,秩约束更是随处可见,但是近年低秩模型的兴起却是受取得了巨大成功的稀疏表示 (Sparse Representation) 和压缩传感 (Compressed Sensing) 理论的推动,由此系统地发展出了新的理论与应用。在此背景下,秩被阐释为二阶 (即矩阵) 稀疏性¹的度量,而不仅仅是一个数学概念。为了说明这一点,我们举图像或视频压缩为例,要实现有效地压缩,必须充分利用图像或视频的时间或空间相关性; 又如 Netflix 挑战² (图 1), 要推断未知的用户评价, 需要充分考虑用户喜好的相关性和视频类别的相关性。矩阵行列间的相关性天然地和矩阵的秩关联在一起。因此,把秩定义为二阶稀疏性度量是很自然的。

下面以我们的工作为基础,简要介绍这方面的进展。我们先介绍线性模型,再介绍非线性模型,然后是常用求解算法,接下来是代表性应用,最后总结; 其中线性模型分单子空间模型和多子空间模型,优化算法分凸优化和非凸优化。

* 本文得到国家自然科学基金(61272341, 61231002)资助。

¹ 一阶稀疏性就是向量的稀疏性,其度量为非零元的个数,即 l_0 范数 $\|\bullet\|_0$ 。

² Netflix 是一家视频租赁公司,拥有很多用户对视频的评价,但这个用户/视频评价矩阵非常稀疏。该公司提供 100 万美元奖金希望能够把预测用户对视频的评价的准确率提高 10%,以便有针对性地推荐,从而提高营收。见 <http://www.netflixprize.com/>

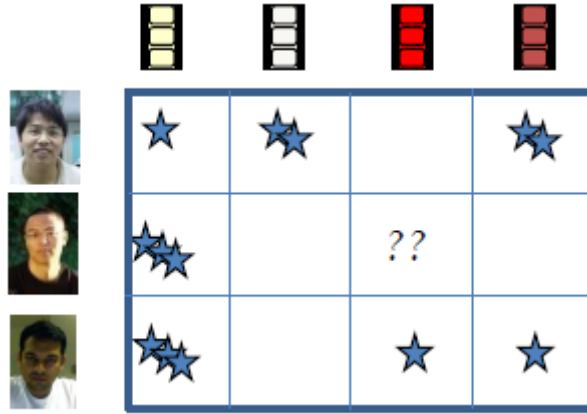


图 1 Netflix 挑战。需要预测用户对其未评价过的视频的喜好程度。

2. 线性模型

近年低秩模型的兴起大致开始于 E. Candès 2008 年提出的矩阵填充 (Matrix Completion, MC) 问题[CR2009]。我们先介绍线性模型。虽然看上去比较简单, 但是理论分析表明线性模型对强噪声和缺失数据非常鲁棒, 在应用中其实也具有足够的表达能力。

a) 单子空间模型

E. Candès 2008 年提出的 MC 问题[CR2009]是: 已知某矩阵 D 在某些位置的值, 可否恢复出该矩阵? 这是一个广泛的数学模型, 如上面提到的 Netflix 挑战、基因微阵列测量等。显然这个问题的答案是不确定的, 鉴于上面提到的需要考虑矩阵行列间的相关性, 他建议选秩最小的那个解 A :

$$\min_A \text{rank}(A), \quad s.t. \quad \pi_\Omega(D) = \pi_\Omega(A), \quad (1)$$

其中 Ω 是已知值的矩阵元素的位置的集合, π_Ω 是保留位置在 Ω 里的矩阵元素的值、其他位置填 0 的投影算子。他考虑的是数据缺失时如何恢复低秩结构的问题。稍后, E. Candès 又进一步考虑了带噪声的 MC 问题[CP2010]:

$$\min_A \text{rank}(A), \quad s.t. \quad \|\pi_\Omega(D) - \pi_\Omega(A)\|_F^2 \leq \varepsilon, \quad (2)$$

以处理测量数据有噪声的情况。

如果考虑数据有强噪声时如何恢复低秩结构的问题, 看似这个问题可以用传统的 PCA 解决, 但实际上传统 PCA 只在噪声是高斯噪声时可以准确恢复潜在的低秩结构。对于非高斯噪声, 如果噪声很强, 即使是极少数的噪声, 也会使传统的主元分析失败。由于主元分析在应用上的极端重要性, 大量学者付出了很多努力在提高主元分析的鲁棒性上, 提出了许多号称“鲁棒”的主元分析方法, 但是没有一个方法被理论上严格证明是能够在一定条件下一定能够精确恢复出低秩结构的。2009 年, Chandrasekaran 等人[CSPW2009]和 Wright 等人[WGRM2009]同时提出了鲁棒主元分析 (Robust PCA, RPCA)。他们考虑的是数据中有稀疏大噪声时如何恢复数据的低秩结构:

$$\min_{A,E} \text{rank}(A) + \lambda \|E\|_0, \quad s.t. \quad D = A + E, \quad (3)$$

其中 $\|E\|_0$ 表示 E 中非零元的个数。J. Wright 的工作后来得到 E. Candès 的加入，获得了更强的结果，即观测矩阵 D 可以只在部分位置知道值。推广后的模型为[CLMW2011]:

$$\min_{A,E} \text{rank}(A) + \lambda \|E\|_0, \quad s.t. \quad \pi_\Omega(D) = \pi_\Omega(A + E). \quad (4)$$

在他们的论文里，也讨论了带稠密高斯噪声的广义 RPCA 模型[CLMW2011]:

$$\min_{A,E} \text{rank}(A) + \lambda \|E\|_0, \quad s.t. \quad \|\pi_\Omega(D) - \pi_\Omega(A + E)\|_F^2 \leq \varepsilon. \quad (5)$$

Chen 等人考虑了噪声集中在若干列的情况，提出了 Outlier Pursuit 模型[CXCS2011]，它把 RPCA 模型中的 $\|E\|_0$ 换成 $\|E\|_{2,0}$ ，即计算 E 的列向量的 l_2 范数有多少个为 0。

Liu 等人把矩阵填充推广到了张量填充[LMWY2013]。虽然张量有基于 CP 分解定义的秩，但是它不可计算，所以 Liu 等人提出了一种折衷的张量秩，定义为张量按不同模式(mode)展开(unfolding)后得到的不同矩阵的秩的和。他们提出的张量填充模型就是：在已知张量某些位置的值时，通过极小化折衷的张量秩来恢复缺失的值。同样基于折衷的张量秩，Tan 等人把 RPCA 推广到了张量恢复[TFFWZ2013]，即把给定张量分解为两个张量之和，一个具有较低的折衷的张量秩，另一个稀疏。

b) 多子空间模型

RPCA 只能从数据中提取一个子空间，它对数据在此子空间中的精细结构无法刻画。精细结构的最简单情形是多子空间模型，即数据分布在若干子空间附近，我们需要找到这些子空间。这个问题马毅等人称为 Generalized PCA (GPCA)问题[VMS2015]，之前已有很多算法，如代数法、RANSAC 等，但都没有理论保障。稀疏表示的出现为这个问题提供了新的思路。E. Elhamifar 和 R. Vidal 2009 年利用样本间相互表达，在表达系数矩阵稀疏的目标下提出了 Sparse Subspace Clustering (SSC)模型[EV2009] ((6)中 $\text{rank}(Z)$ 换成 $\|Z\|_0$ ，同时添加约束 $\text{diag}(Z) = 0$ 以防止只用样本本身表达自己)。受此启发，刘光灿等人提出了 Low-Rank Representation (LRR)模型[LLY2010][LLYSYM2013]:

$$\min_{Z,E} \text{rank}(Z) + \lambda \|E\|_0, \quad s.t. \quad D = DZ + E. \quad (6)$$

之所以要求表达系数矩阵 Z 低秩，是为了增强 Z 各列之间的相关性以提高对噪声的抵抗能力。SSC 和 LRR 的最优表达系数矩阵 Z^* 可以作为样本间的相似性度量，用 $(\|Z^*\|_1 + \|Z^{*T}\|_1) / 2^3$ 定义样本间的权重 ($\|Z^*\|_1$ 表示把 Z^* 的元素都取绝对值得到的矩阵)，再通过谱聚类就可以把数据聚类成若干线性子空间。由于最优表达系数矩阵 Z^* 可以作为样本间的相似性度量，庄连生等人进一步要求系数矩阵稀疏、非负，以应用于半监督学习[ZGLMZY2012]。

LRR 需要假定数据充足。在样本不足情形，刘光灿和颜水成[LY2011]提出了 Latent LRR 模型:

³ 在后来的 TPAMI2013 论文[LLYSYM2013]中，刘光灿等人改用 $\left|U_{Z^*} U_{Z^*}^T\right|$ 构建邻接矩阵，其中 U_{Z^*} 是 Z^* 的瘦型奇异值分解的左奇异向量集合。原因见第 2.C.i 节。

$$\min_{Z,L,E} \text{rank}(Z) + \text{rank}(L) + \lambda \|E\|_0, \quad s.t. \quad D = DZ + LD + E. \quad (7)$$

他们称 DZ 为 Principal Feature, LD 为 Salient Feature, Z 用于子空间聚类, L 则可用于提取数据的鉴别信息以识别。而刘日升等人[LLTS2012]提出了 Fixed Rank Representation (FRR)模型:

$$\min_{Z,\tilde{Z},E} \|Z - \tilde{Z}\|_F^2 + \lambda \|E\|_{2,0}, \quad s.t. \quad D = DZ + E, \text{rank}(\tilde{Z}) \leq r, \quad (8)$$

其中 \tilde{Z} 用于度量样本间的相似性。

为了进一步提高子空间聚类的精度, 卢参义等人提出使用 Trace Lasso 来约束表达系数 [LFLY2013]:

$$\min_{z_i, e_i} \|D \text{diag}(z_i)\|_* + \lambda \|e_i\|_0, \quad s.t. \quad d_i = Dz_i + e_i, \quad i = 1, \dots, n, \quad (9)$$

其中 z_i 是矩阵 Z 的第 i 列, $\|D \text{diag}(z_i)\|_*$ 称为向量 z_i 的 Trace Lasso, $\|\cdot\|_*$ 为矩阵的核范数 (Nuclear Norm, 即矩阵奇异值之和)。当 D 的列是 l_2 -范数归一化的时, Trace Lasso 具有优美的插值性质:

$$\|z_i\|_2 \leq \|D \text{diag}(z_i)\|_* \leq \|z_i\|_1,$$

而且左端等式在数据是完全相关时 (每列相同或方向相反) 达到, 右端在数据是完全无关时 (列之间正交) 达到。因此, Trace Lasso 具有自适应于数据相关性的特性。该模型称为 Correlation Adaptive Subspace Segmentation (CASS)。

为了更有效地对张量数据进行聚类, 傅逸凡等人提出了 Tensor LRR 模型[FGTL2014], 以便综合张量在各模式下的信息。

c) 理论分析

对低秩模型的理论分析是比较丰富的, 可以归纳为以下三个部分。

i. 精确恢复

上面介绍的低秩模型都是离散优化问题, 很多都是 NP-难, 这对求解它们造成了很大的困难。为了克服这个困难, 一个常用的办法是把它们近似成凸优化问题。粗略地说, 和 l_0 范数 $\|\cdot\|_0$ 最接近的凸函数是 l_1 范数 $\|\cdot\|_1$, 即所有元素绝对值的和; 和秩最接近的凸函数是核范数 $\|\cdot\|_*$ 。这样, 上面的

离散优化问题都可以转化为凸优化问题，可以比较方便地求解。但是会自然产生一个问题：求解凸优化问题是否能够得到真实的解？大部分针对单子空间的低秩模型，如 MC[CR2009]、RPCA[CLMW2011]、带缺失数据的 RPCA[CLMW2011]、Outlier Pursuit[CXCS2011][ZLZC2015]，都有相应的肯定回答，基本上都可以归纳为：如果 Outlier 是稀疏的且随机分布的、真实解矩阵是低秩的，则真实解矩阵是可以精确恢复的。需要指出的神奇之处是：精确恢复性与 Outlier 的大小无关，而是与其稀疏程度有关。这就保证了大部分针对单子空间的低秩模型是具有很强的鲁棒性的。这个特性和传统的 PCA 的差别是非常显著的。而对于针对多子空间的低秩模型，只有 LRR 有较为详细的理论分析[LXY2012]，但刘光灿等人仅证明了当 Outlier 比例不超过某阈值时， Z_0 的行空间和哪些样本是 Outlier 是可以精确恢复的， Z_0 的行空间可以从 $U_{Z^*} U_{Z^*}^T$ 得到，其中 $U_{Z^*} \Sigma_{Z^*} V_{Z^*}^T$ 为最优解 Z^* 的瘦型奇异值分解。该分析没有回答 Z_0 和 E_0 本身是否可以精确恢复，但是幸运的是，将 LRR 应用于子空间聚类时，我们只需要 Z_0 的行空间就够了。

ii. 闭解

低秩模型一个非常令人振奋的性质在于它在无噪声情形可能有闭解，这是稀疏模型所不具备的性质。魏嗣明和林宙辰[WL2010]分析了 LRR 的数学性质，首先发现无噪声的 LRR 模型：

$$\min_Z \|Z\|_*, \quad s.t. \quad D = DZ, \quad (10)$$

有唯一解，而且该解可以很方便地表达出来：设 D 的瘦型奇异值分解为 $U_D \Sigma_D V_D^T$ ，则该解为 $V_D V_D^T$ 。

矩阵 $V_D V_D^T$ 在立体视觉里被称为 Shape Interaction Matrix。刘光灿等人[LLYSYM2013]进一步发现使用一般字典的 LRR 模型：

$$\min_Z \|Z\|_*, \quad s.t. \quad D = BZ, \quad (11)$$

也有唯一解，而且该解也可以很方便地表达出来： $Z^* = B^+ D$ ，其中 B^+ 为 B 的 Moore-Penrose 伪逆。

这个结论被 Yu 和 Schuurmans 推广到一般的正交不变范数[YS2011]，在该文中作者找到了更多的具有闭解的低秩问题。Favaro 等人也找到了若干和子空间聚类有关的带闭解的低秩问题[FVR2011]。张宏扬等人进一步发现无噪声的 Latent LRR 问题（离散形式和凸形式）的解不唯一，并给出了所有闭解[ZLZ2013]，文中还发现离散形式的无噪声 LRR 问题（即(6)中 E 为 0）事实上不是 NP-难的，并给出了所有闭解。张宏扬等人因此提出从 Latent LRR 解集中寻找最稀疏解的算法来弥补 Latent LRR 的这个缺陷[ZLZG2014]。

iii. 块对角结构

面向多子空间的低秩模型都会算得一个表达系数矩阵 Z 。对于 SRC 和 LRR，都可以证明：在理想情况下，即当样本无噪声、子空间相互独立（即任何子空间都不能用其他子空间来表达）时，

最优的表达系数矩阵 Z^* 是块对角的。由于每个对角块对应于一个子空间， Z^* 的块对角结构对子空间聚类至关重要。令人惊讶的是，卢参义等人证明了：如果 Z 使用 Frobenius 范数的平方来正则化（相应模型称为 LSR），则在理想情况下最优表达系数矩阵 Z^* 也是块对角的[LZZHY2012]。为此，卢参义等人提出了强化块对角条件（Enforced Block-Diagonal Conditions），只要对 Z 的正则化项满足强化块对角条件，则在理想情况下最优解就是块对角的[LZZHY2012]，这就大大拓宽了 Z 的正则化项的选择范围，即可以不限是稀疏或低秩约束。对于 Z 按列求解的子空间聚类模型，如基于 Trace Lasso 的 CASS 模型(9)，卢参义等人也相应提出了强化块稀疏条件（Enforced Block-Sparse Conditions），只要对 Z 的列的正则化项满足强化块稀疏条件，则在理想情况下最优的 Z 就是块对角的[LFLY2013]。但是以上结论都是在理想情况下分析的，在有噪声或子空间之间不独立情形，最优的 Z 会偏离块对角，对后续的寻找子空间造成一定的困难。为此，冯佳时等人根据谱图理论关于 Laplacian 矩阵 0 特征值重数和权重矩阵对角块个数之间的对应关系提出了块对角先验[FLXY2014]，把它添加到前面的子空间聚类模型中，就可以保证在任何情况下得到的都是严格块对角的表达系数矩阵 Z ，从而显著提高了对噪声的抵抗能力。

表达系数之间的群组效应（Grouping Effect），即当样本相近时其表达系数向量也相近，对在噪声情况下保持表达系数矩阵 Z 的块对角结构有一定作用。SSC、LRR、LSR、CASS 等都证明了具有群组效应。胡瀚等人为此提出了一般性的强化群组效应条件（Enforced Grouping Effect Conditions）[HLFZ2014]，可以很方便地判定正则化函数具有群组效应。

3. 非线性模型

用于分割非线性流形的低秩模型比较少。一个自然的想法是利用 Kernel 技巧，由 Wang 等人提出[WSC2011]。其思想是：假设通过非线性映射 ϕ ，样本集 X 在高维空间中分布在线性子空间上，则可以对映射后的样本集运用 LRR 模型。假定噪声是高斯的，则模型为：

$$\min_Z \|\phi(X) - \phi(X)Z\|_F^2 + \lambda \|Z\|_*$$

由于 $\|\phi(X) - \phi(X)Z\|_F^2 = \text{tr}[(\phi(X) - \phi(X)Z)^T (\phi(X) - \phi(X)Z)]$ ，就得到了 $\phi^T(X)\phi(X)$ ，就可以引入核函数 $K(x, y)$ ，使得 $K(x, y) = \phi^T(x)\phi(y)$ 。因此，上面的模型可以写成核化的形式而不需显示引入非线性映射 ϕ 。但是，当噪声不是高斯时，上述核技巧不适用。

4. 优化算法

第 2 节中所列离散低秩模型一般都是 NP-难的，只能近似求解。一般的做法是转化成连续优化问题。这时有两种转化方式，一种是转化为凸优化问题，如上面提到的，把 l_0 范数 $\|\cdot\|_0$ 换成 l_1 范数 $\|\cdot\|_1$ 、

秩换成核范数 $\|\cdot\|_*$ 。另一种是转化为非凸优化问题，就是用非凸的连续函数来近似 l_0 范数 $\|\cdot\|_0$

（比如用 l_p 范数 $\|\cdot\|_p$ ($0 < p < 1$) 和秩（比如用 Schatten- p 范数（奇异值的 l_p 范数））。还有一种做法是把需要约束为低秩的矩阵直接表达成两个矩阵的乘积，第一个矩阵的列数和第二个矩阵的行数均为期望的秩，然后交替更新第一个和第二个矩阵直至不变为止。这种做法是在稀疏模型中所没有的。凸优化的好处是能够得到修改后的模型的全局最优解，但是解有可能不够低秩或稀疏；非凸优化的好处是往往能得到更低秩和更稀疏的解，但是不能得到修改后的模型的全局最优解，解的质量可能依赖于初值。所以凸优化和非凸优化互为补充。针对问题的特点，还有可能设计随机算法，大大降低求解的复杂度。

a) 凸优化算法

凸优化已经是比较成熟的领域，一般都有多项式时间复杂度的求解算法（比如内点法）。对于大规模/高维数据，一般需要 $O(n \text{poly} \log(n))$ 的复杂度，连平方复杂度都是不可容忍的。为了帮助解低秩模型所面临的计算复杂度的障碍，我们举以下例子说明。对于 RPCA 问题，如果矩阵大小是 $n \times n$ ，则该问题有 $2n^2$ 个未知数，即使 $n = 1000$ ，它对应于一个不大的矩阵，未知数个数也达到了一百万。如果使用内点法来求解，比如使用斯坦福大学的 CVX 软件包，则每迭代一次的时间复杂度是 $O(n^6)$ ，而空间复杂度则是 $O(n^4)$ ，导致一般内存为 4GB 的 PC 机只能处理 80×80 的矩阵。因此要使得低秩模型实用，必须要设计高效的优化算法。

目前面向大规模计算的优化算法都是一阶算法。代表性的算法包括加速近邻梯度法（Accelerated Proximal Gradient, APG）[Nest1983][BT2009]和交错方向法（Alternating Direction Method, ADM）[LCM2009][LLS2011][LLL2015]。APG 主要针对无约束问题。在目标函数为 $C^{1,1}$ ，即可微且梯度 Lipschitz 连续时，一般的梯度下降法的收敛速度只能达到 $O(k^{-1})$ ，其中 k 为迭代次数，但是 Nesterov 构造了一个算法[Nest1983]，其收敛速度可以达到 $O(k^{-2})$ 。后来 Beck 和 Teboulle 把 Nesterov 的算法推广到目标函数为一个凸函数和一个 $C^{1,1}$ 函数和的情形[BT2009]，大大扩展了 Nesterov 算法的应用范围。另外，APG 需要估计 Lipschitz 系数，如果 Lipschitz 系数估计得过于保守（太大），则会影响收敛速度，所以 Beck 和 Teboulle 还进一步提出了使用动态估计的 Lipschitz 系数，以加速收敛[BT2009]。在一些有特殊结构的问题上，APG 可以被推广（Generalized APG, GAPG）[ZL2011]，使得不同变量对应的 Lipschitz 系数不同，从而加快收敛。对于带线性约束的优化问题，可以把约束的平方放到目标函数里作为惩罚项，再应用 APG 求近似解。为加快收敛速度，惩罚系数要逐渐增加到较大的值，这一重要技巧称为 Continuation[GLWWCM2009]。ADM 面向目标函数可分解（即为两个不同变量凸函数的和）、带线性约束和凸集约束的问题，它是 Lagrange 乘子法的一种变形。它先构造问题的增广 Lagrangian 函数，然后通过交替极小化增广 Lagrangian 函数来更新两个变量，最后更新 Lagrange 乘子，如此迭代[LCM2009]。其好处是更新变量子问题要比原问题简单，甚至有闭解。在子问题不容易求解的情形，可以考虑线性化增广 Lagrangian 函数的增广项，把问题进一步简化，这一技术称为线性化交错方向法（Linearized Alternating Direction Method, LADM）[LLS2011]。如果

线性化增广 Lagrangian 函数的增广项还不足以导致足够简单的子问题，还可以进一步把目标函数的 $C^{1,1}$ 成分也线性化[LLL2015]。对于多变量（变量数大于 2）的凸规划问题，简单地套用两个变量的 ADM 法不能保证收敛。但是如果把串行更新改成并行更新，通过适当的参数选取，则可以保证收敛，带线性化的情形也是如此[LLL2015]。在以上介绍的 ADM 系列算法中，增广项的惩罚参数可以动态增长，以加速收敛[LLS2011][LLL2015]。

无论是用何种凸优化算法求解低秩模型，都会碰到如下形式的子问题：

$$\min_X \left\{ \|X\|_* + \frac{\alpha}{2} \|X - W\|_F^2 \right\}.$$

它有闭解[CCS2010]：设 W 的奇异值分解为 $W = U\Sigma V^T$ ，则最优解为 $X = U\Theta_{\alpha^{-1}}(\Sigma)V^T$ ，其中

$$\Theta_{\varepsilon}(x) = \begin{cases} x - \varepsilon, & \text{如果 } x > \varepsilon, \\ x + \varepsilon, & \text{如果 } x < -\varepsilon, \\ 0, & \text{如果 } -\varepsilon \leq x \leq \varepsilon. \end{cases} \quad (12)$$

因此求解低秩模型往往离不开奇异值分解，而维度为 $m \times n$ 的矩阵的奇异值分解的复杂度为 $O(mn \min(m, n))$ ，所以一般来说求解低秩模型计算量都很大，在面临大矩阵时问题尤为突出。所幸从(12)可以看出小于 α^{-1} 的奇异值及其奇异向量无需求出，因为这些奇异值将被收缩为 0，从而对

X 没有贡献。于是就可以只计算大于 α^{-1} 的奇异值及其奇异向量，这可以使用 PROPACK 来实现，

计算量相应地下降到 $O(rmn)$ ，其中 r 为最优 Z 的秩。值得一提的是，PROPACK 只能提供指定个数的最大奇异值及其奇异向量，所以调用 PROPACK 时要动态地预测 r 的值[LCM2009]。当解不是低秩的时，比如在图像处理和计算机视觉里已经有广泛应用的 Transform Invariant Low-Rank Textures (TILT)[ZGLM2012]（见(13)及第 5.3 节），可以使用增量奇异值分解[RL2013]进行加速。

b) 非凸优化算法

对于用 Schatten- p 范数近似秩、 l_p 范数近似 l_0 范数的无约束问题，一个有效的做法是迭代重加权最小二乘法（Iteratively Reweighted Least Squares）[LLY2015]，即把 $tr\left(\left(X\tilde{X}\right)^{\frac{p}{2}}\right)$ 近似为

$$tr\left(\left(X_k X_k^T\right)^{\frac{p-1}{2}}\left(XX^T\right)\right), |x_i|^p \text{ 近似为 } |x_i^{(k)}|^{p-2} x_i^2, \text{ 其中 } X_k \text{ 为低秩矩阵 } X \text{ 在第 } k \text{ 次迭代的值, } x_i^{(k)}$$

为稀疏向量 x 的第 i 个分量在第 k 次迭代的值。这样每次更新 X 需要解矩阵方程，更新 x 需要求解线性方程组。另一种方法是借用 APG 的思想，把目标函数的 $C^{1,1}$ 成分也线性化，这样每次只需求解如下形式的子问题：

$$\min_X \left\{ \sum_{i=1}^n g(\sigma_i(X)) + \frac{\alpha}{2} \|X - W\|_F^2 \right\},$$

其中 g 是单调递增的凹函数，比如 x^p ($0 < p < 1$)。卢参义等人给出了求解上面问题的算法 [LZXYL2015]。

另一种近似秩的函数是截断核范数 (Truncated Nuclear Norm, TNN) [HZYLH2013]:

$$\|X\|_r = \sum_{i=r+1}^{\min(m,n)} \sigma_i(X),$$

它不包含前 r 个奇异值，因此不是凸函数。可以很容易看出来，极小化 TNN

会促使后面的奇异值尽可能小，而前 r 个奇异值取值多少没有影响，因此可以促使解更接近一个秩 r 的矩阵。TNN 可以更进一步，给越往后的奇异值加越大的权重，得到加权核范数 (Weighted Nuclear

$$\text{Norm, WNN}) [GZZF2014]: \|X\|_{w,*} = \sum_{i=1}^{\min(m,n)} w_i \sigma_i(X).$$

但此时求解子问题:

$$\min_X \left\{ \|X\|_{w,*} + \frac{\alpha}{2} \|X - W\|_F^2 \right\},$$

一般没有闭解，需要解一个小型的凸优化问题。

第三种方法是直接把待求的低秩矩阵 X 表达成 $X = AB^T$ ，其中 A 和 B 都是列数为 r 的矩阵，则可以交替更新 A 和 B 直至不变 [WYZ2012]。这种方法的优势在于简单，缺点在于需要较好地先验估计最优解的秩。

c) 随机优化算法

以上方法，无论凸或是非凸，计算复杂度都至少是 $O(rmn)$ ，其中 m 、 n 是待求低秩矩阵的尺寸，这在 m 、 n 都很大时还是不够快。要突破这个复杂度瓶颈，需要使用随机算法。但是随机算法不能简单地把确定型算法的每一步随机化，因为有些步骤的随机化会导致很大的误差。所以随机算法需要根据低秩模型的特点来设计，导致目前这方面工作还不多。对 RPCA，刘日升等人提出了 l_1 -滤波方法 [LLSG2014]，先随机采样数据矩阵 D 的适当大小的子矩阵 D^s ， D^s 上求解小型 RPCA 问题得到地址的 A^s 和稀疏的 E^s ，然后 D^s 所在的其他行和列用求得的 A^s 来处理，最后原数据矩阵 D 对应的低秩矩阵 A 可以通过 Nystrom 技巧来表出，整个算法的复杂度为 $O(r^3) + O(r^2(m+n))$ ，为关于矩阵大小线性。对 LRR 和 Latent LRR，张宏扬等人发现如果模型略作改动，其解和 RPCA 的解可

以互为表出，因此，都可以通过先求解 RPCA 来大大加速[ZLZG2015]。

5. 代表性应用

低秩模型在信号处理和机器学习等领域里已经获得了广泛的应用，NIPS2011 上曾出现了大量的讨论低秩模型的论文。由于专业所限，本节只简要介绍笔者及其合作者在图像处理和计算机视觉领域里找到的典型应用。

5.1 背景建模[CLMW2011]

背景建模的最简单情形是从固定摄像机拍摄的视频中分离背景和前景。此时很容易想到背景是基本不变的，所以如果把背景的每一帧作为矩阵的一列，则该矩阵低秩。同时由于前景是移动的物体，占据像素比例较低，所以前景对应于视频中的稀疏“噪声”部分。由此得到做背景建模的 RPCA 模型(3)，其中 D 的每一列是视频的每一帧拉直后得到的向量， A 的每一列对应于背景的每一帧拉直后得到的向量， E 的每一列对应于前景的每一帧拉直后得到的向量。部分结果如下：

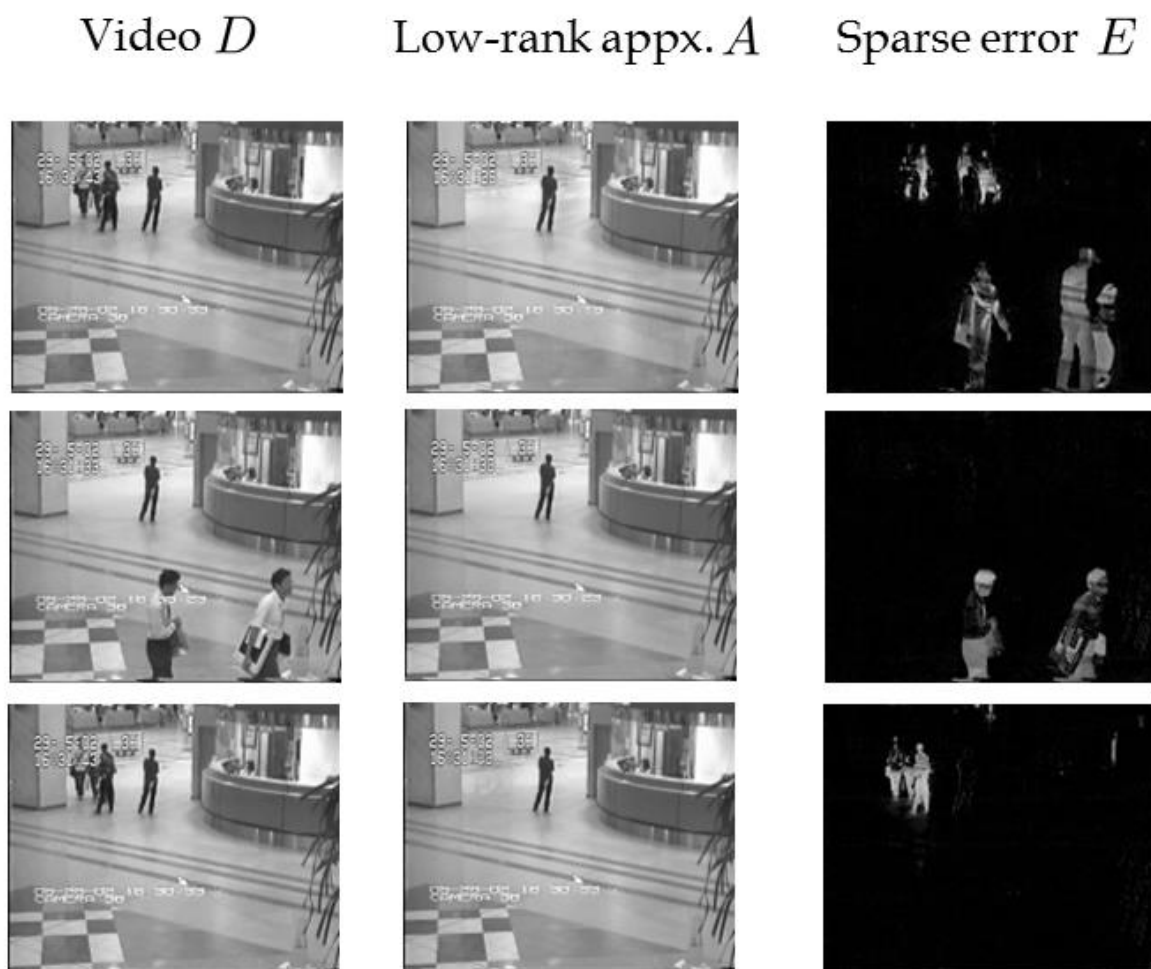


图 2 背景建模。第一列是监控视频，第二列是背景视频，第三列是前景视频（ E 值取绝对值）。

5.2 图像批量对齐 (RASL) [PGWXM2012]

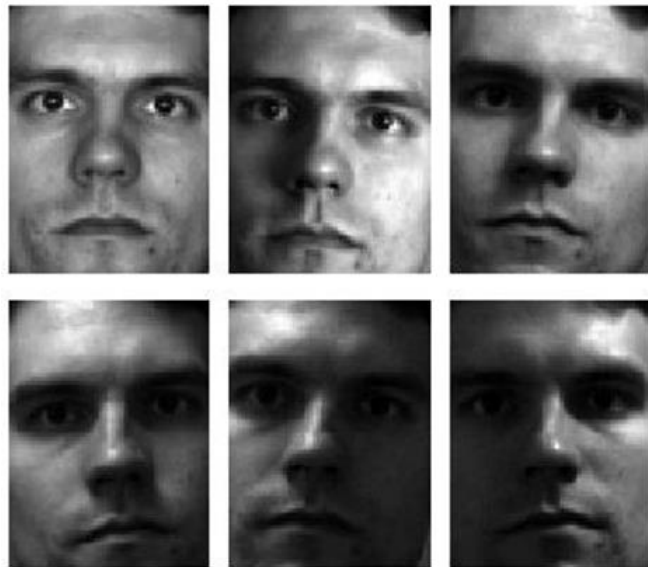
背景建模需要假定背景已经对齐，这样才能得到低秩的背景视频。在没有对齐的情况下，可以考虑把每一帧/每一幅图像作适当的几何变形使它们对齐。此时数学模型为：

$$\min_{\tau, A, E} \|A\|_* + \lambda \|E\|_1, \quad s.t. \quad D \circ \tau = A + E, \quad (13)$$

其中 $D \circ \tau$ 是把每一帧/每一幅图像作适当的几何变形 τ 的形式写法（每一帧/每一幅图像的几何变形不同）。此时，(35) 是一个非凸优化问题。为了有效求解，彭义刚等人[PGWXM2012]提出了把 τ 局部线性化的迭代算法，即做如下循环直至 $\Delta\tau_k$ 足够小：

$$\begin{cases} \min_{\Delta\tau_k, A, E} \|A\|_* + \lambda \|E\|_1, & s.t. \quad D \circ \tau_k + J\Delta\tau_k = A + E, \\ \tau_{k+1} \leftarrow \tau_k + \Delta\tau_k, \\ k \leftarrow k + 1, \end{cases} \quad (14)$$

其中 J 是 $D \circ \tau$ 对 τ 的参数的 Jacobi 矩阵。在仿射变换模型下，人脸图像的部分对齐结果如下：



人脸的初始姿态

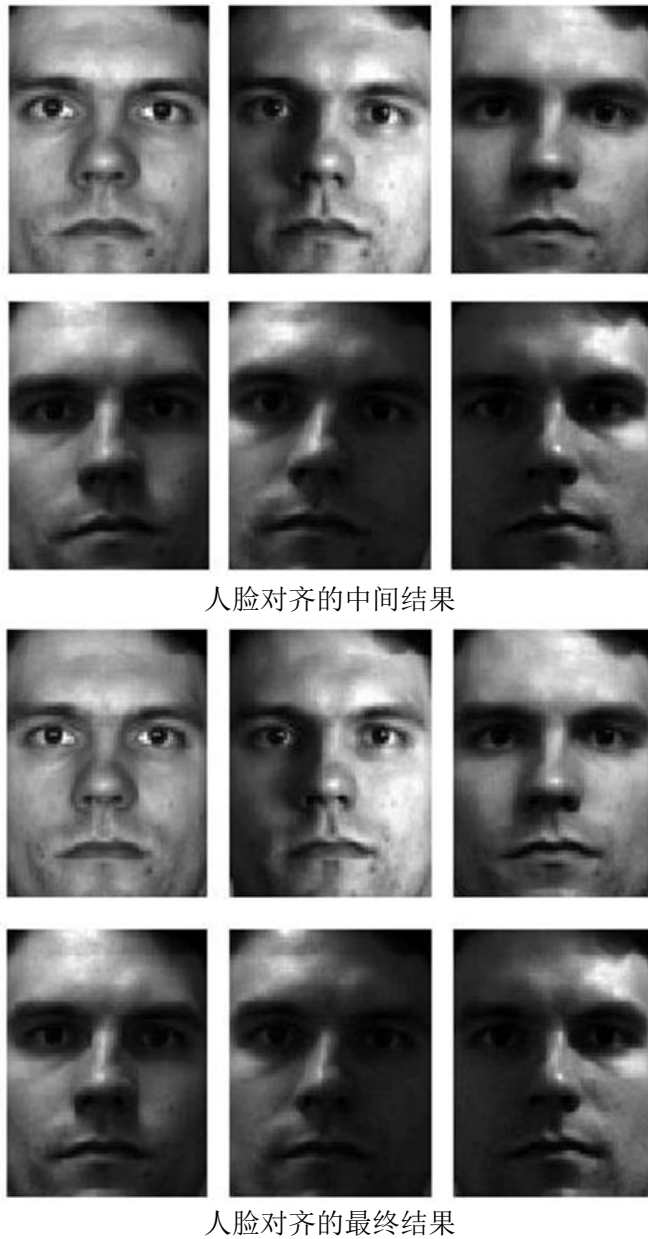


图 3 人脸对齐的迭代过程

5.3 变换不变低秩纹理 (TILT) [ZGLM2012]

变换不变低秩纹理 (Transform Invariant Low-rank Textures, TILT) 的数学模型形式上和 RASL 的 (35) 完全一样, 其求解过程也完全一样, 但是这里 D 是图像的一个长方形图像块。TILT 的思想是通过几何变换 τ 把 D 所代表的图像区域校正成正则的区域, 如具有横平竖直、对称等特性, 这些特性可以通过低秩性来进行刻画。以下是射影变换下图像校正的例子:



图 4 利用 TILT 模型进行图像校正的例子。第一行表示原始的图像块（长方形框）及相应校正变换（四边形框，校正变换就是四边形框相对于长方形框的几何变换），第二行是校正后的图像块。

TILT原则上对任何参数变换都适用。张铮东等人[ZLM2011]还考虑了基于广义柱体变换的TILT，可用于人造建筑物表面纹理的提取。一些例子如下：

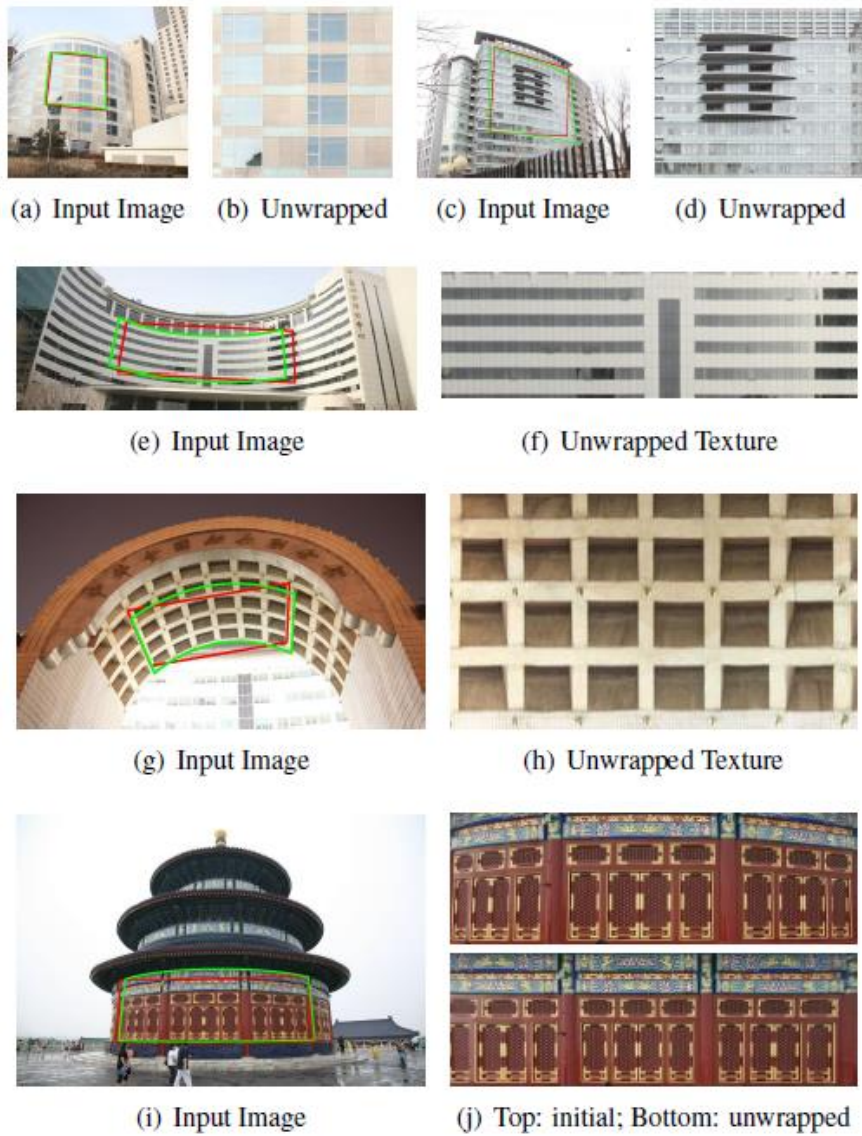


图 5 基于广义柱体变换的 TILT 的建筑物纹理展开

TILT 还被广泛应用于建筑物几何建模[ZGLM2012]、相机自动标定和镜头畸变自动校正[ZMM2011]、文字校正[ZLSM2013]与检测[ZLSM2014]等。由于其应用上的重要性，任翔和林宙辰[RL2013]特别研究了它的基于 LADM 的快速算法，把求解速度提高了 5 倍以上。

5.4 运动分割[LLY2010, LLYSYM2013]

LRR 被认为是做刚体运动分割最好的算法之一[AEH2013]。所谓刚体运动分割，就是把视频里做刚体运动的物体上的特征点进行聚类，使得每一类对应于一个独立运动的物体，这样就可以得到物体运动的轨迹。部分例子如下：



图 6 运动分割的例子

5.5 图像分割[CLWHY2011]

图像分割是特殊的聚类问题。首先把图像过分割（Over Segment）成超像素（Super-pixel），然后在超像素上提取适当的特征，通过改进的 LRR 模型综合多种特征（基本上每一种特征对应于一个 LRR 模型），求出整体表示矩阵 Z^* ，然后对用 $(|Z^*| + |(Z^*)^T|) / 2$ 表出的相似性矩阵进行正则化割（Normalized Cut），得出超像素的聚类关系，每一类就对应于一个图像区域。部分例子如下：

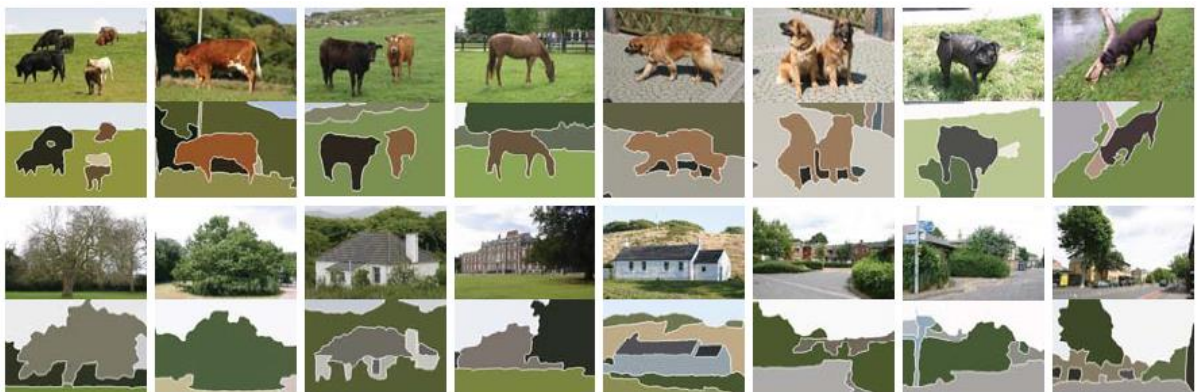


图 7 图像分割的例子

5.6 图像显著区域检测[LLYY2012]

运动分割和图像分割都是利用了 LRR 的表示矩阵 Z ，而图像显著区域检测则是利用 LRR 里的稀疏“噪声” E 。图像显著区域一般就是图像中“与众不同”的区域，因此如果用其他区域进行“预测”则会产生较大的误差。因此，如果把图像分解成小块，在其上提取适当特征，则图像显著区域对应于 LRR 里的稀疏“噪声” E 较大的部分。部分例子如下：

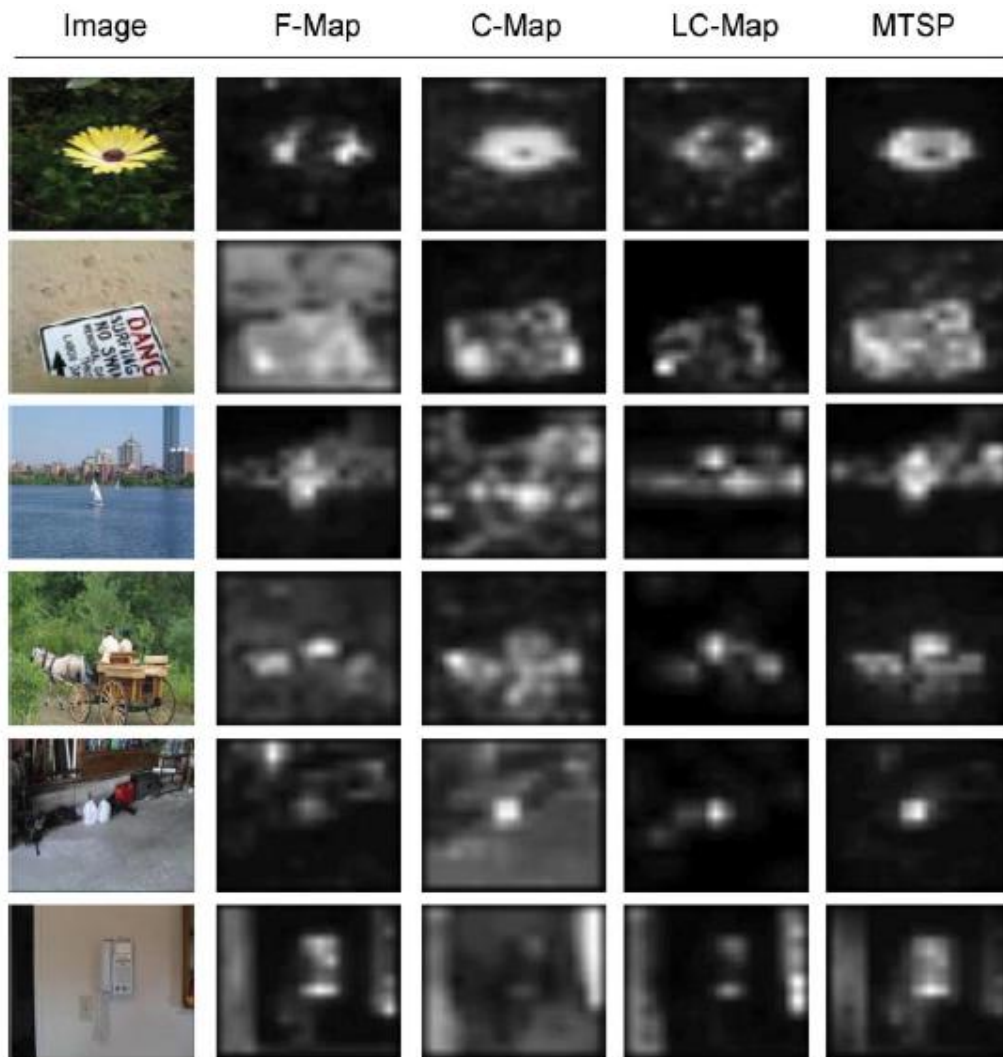


图 8 图像显著区域检测的例子。第一列是输入图像，第二至五列是不同方法的检测结果，其中最后一列是基于 LRR 的检测方法。

5.7 其他应用

低秩模型的其他应用，如部分重复图像检索[YLLZ2014]、人脸识别[QYZL2014]、结构化纹理修复、街景全景图拼接、人造物体定向(Upright Orientation)、基于光度学的立体视觉[WGSMWM2010]、图像标签的改进[ZYM2010]、视觉域适应(Visual Domain Adaption)[JLLC2012]、鲁棒视觉跟踪[ZGLA2012]、三维人脸特征提取[MR2012]、CT重建[GCSZ2011]、图像半监督分类[ZGLMZY2012]、图像集的协同分割(co-segmentation)[MSXC2012]、视点无关步态识别、文档关键词提取、甚至音频分析[PK2012]、蛋白质-基因相关性分析、网络流量异常检测、鲁棒滤波与系统辨识等等，限于篇幅就不再一一介绍了。

6. 结束语

低秩模型近年来信号处理、机器学习、计算机视觉等领域获得了广泛的应用，短短几年从理论、算法到应用各方面都得到了快速的发展，本文只是基于我们的工作做了非常粗浅的介绍。许多具体

问题，如果结合问题的特性适当地引入低秩性约束，很多情况下都能得到更好的结果。有些问题的数据本身可能没有低秩性，此时可以引入适当变换增强其低秩性（如 RASL/TILT 对 RPCA 的改进）。有些学者没有检查数据是否具有低秩性或对数据进行适当的预处理就声称使用低秩约束效果不好，这是不太严谨的。从上面的介绍可以看出，低秩模型在如下几个方面还需要更多的研究：矩阵往张量的推广、非线性流形聚类、近线性复杂度的随机算法等。希望我们的介绍能起到抛砖引玉的作用。

参 考 文 献

- [AEH2013] Adler A, Elad M, Hel-Or Y. Probabilistic subspace clustering via sparse representations. *IEEE Signal Processing Letters*, 2013, 20(1):63 – 66.
- [BT2009] Beck, A, Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2009, 2(1):183-202.
- [CCS2010] Cai J-F, Candès E J, Shen Z. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 2010, 20(4): 1956-1982.
- [CLMW2011] Candès E J, Li X, Ma Y, Wright J: Robust principal component analysis? *Journal of the ACM*, 2011, 58(1): 1-37.
- [CP2010] Candès E J, Plan Y. Matrix completion with noise. *Proceedings of the IEEE*, 2010, 98(6): 925-936.
- [CR2009] Candès E J, Recht B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009, 9(6): 717-772.
- [CSPW2009] Chandrasekaran V, Sanghavi S, Parrilo P, Willsky A. Sparse and low-rank matrix decompositions. In: *Allerton'09 Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, 2009, 962-967.
- [CXCS2011] Chen, Y., Xu, H., Caramanis, C., and Sanghavi, S. Robust matrix completion and corrupted columns. In *International Conference on Machine Learning*, 2011, pp. 873–880.
- [CLWHY2011] Cheng B, Liu G, Wang J, Huang Z, Yan S. Multi-task low-rank affinity pursuit for image segmentation. In: *Proceedings of International Conference on Computer Vision (ICCV)*, 2011, 2439-2446.
- [EV2009] Elhamifar E, Vidal R. Sparse subspace clustering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, vol. 2, 2790-2797.
- [FVR2011] P. Favaro, R. Vidal, and A. Ravichandran. A closed form solution to robust subspace estimation and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1801-1807, 2011.
- [FLXY2014] Jiashi Feng, Zhouchen Lin, Huan Xu, and Shuicheng Yan. Robust Subspace Segmentation with Block-diagonal Prior. *CVPR* 2014.
- [FGTL2014] Yifan Fu, Junbin Gao, David Tien, and Zhouchen Lin, Tensor LRR Based Subspace Clustering, *International Joint Conference on Neural Networks (IJCNN)* 2014.
- [GLWWCM2009] Ganesh A, Lin Z, Wright J, Wu L, Chen M, Ma Y. Fast algorithms for recovering a corrupted low-rank matrix. In: *Proceedings of 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2009, 213-216.
- [GCSZ2011] Gao H, Cai J-F, Shen Z, Zhao H. Robust principal component analysis-based four-dimensional computed tomography. *Physics in Medicine and Biology*, 2011, 56:3181–3198.
- [GZZF2014] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted Nuclear Norm Minimization with Application to Image Denoising, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [HLFZ2014] Han Hu, Zhouchen Lin, Jianjiang Feng, and Jie Zhou. Smooth Representation Clustering. *CVPR* 2014.
- [HZYLH2013] Yao Hu, Debing Zhang, Jieping Ye, Xuelong Li, and Xiaofei He. Fast and Accurate Matrix

- Completion via Truncated Nuclear Norm Regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 9, pp. 2117-2130, 2013.
- [JLLC2012] Jhuo I-H, Liu D, Lee D T, Chang S-F. Robust visual domain adaptation with low-rank reconstruction. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, 2168-2175.
- [LLYY2012] Lang C, Liu G, Yu J, Yan S. Saliency detection by multitask sparsity pursuit. *IEEE Transactions on Image Processing*, 2012, 21(3): 1327-1338.
- [LCM2009] Lin Z, Chen M, Ma Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrix. University of Illinois at Urbana Champaign, Technical Report UILU-ENG-09-2215, October 2009 (arXiv: 1009.5055).
- [LLL2015] Zhouchen Lin, Risheng Liu, and Huan Li. Linearized Alternating Direction Method with Parallel Splitting and Adaptive Penalty for Separable Convex Programs in Machine Learning, *Machine Learning*, 2015.
- [LLS2011] Lin Z, Liu R, Su Z. Linearized alternating direction method with adaptive penalty for low-rank representation. In: *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2011, 612-620.
- [LLSG2014] Risheng Liu, Zhouchen Lin, Zhixun Su, and Junbin Gao, Solving Principal Component Pursuit in Linear Time via l1 Filtering, *Neurocomputing*, Vol. 142, pp. 529-541, 2014.
- [LLTS2012] Risheng Liu, Zhouchen Lin, Fernando De La Torre, and Zhixun Su, Fixed-Rank Representation for Unsupervised Visual Learning, *CVPR 2012*.
- [LLYSYM2013] Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 171-184.
- [LLY2010] Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation. In: *Proceedings of International Conference on Machine Learning (ICML)*, 2010, 663-670
- [LMWY2013] Ji Liu, Przem Musialski, Peter Wonka, and Jieping Ye. Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 1, pp. 208-220, 2013.
- [LXY2012] Liu G, Xu H, Yan S. Exact subspace segmentation and outlier detection by low-rank representation. In: *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTAT)*, 2012, 703-711.
- [LY2011] Guangcan Liu, Shuicheng Yan. Latent Low-Rank Representation for subspace segmentation and feature extraction. *ICCV 2011*.
- [LFLY2013] Can-Yi Lu, Jiashi Feng, Zhouchen Lin, Shuicheng Yan. Correlation Adaptive Subspace Segmentation by Trace Lasso. *ICCV 2013*: 1345-1352.
- [LLY2015] Canyi Lu, Zhouchen Lin, and Shuicheng Yan. Smoothed Low Rank and Sparse Matrix Recovery by Iteratively Reweighted Least Squared Minimization, *IEEE Trans. Image Processing*, 2015.
- [LZZHY2012] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, Shuicheng Yan. Robust and Efficient Subspace Segmentation via Least Squares Regression. *ECCV (7) 2012*: 347-360.
- [LZXYL2015] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized Singular Value Thresholding. *AAAI 2015*.
- [MR2012] Ming Y, Ruan Q. Robust sparse bounding sphere for 3D face recognition. *Image and Vision Computing*, 2012, 30:524-534.
- [MSXC2012] Mukherjee L, Singh V, Xu J, Collins M D. Analyzing the subspace structure of related images: Concurrent segmentation of image sets. In: *Proceedings of European Conference on Computer Vision (ECCV)*, 2012, 128-142.

- [Nest1983] Nesterov, Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Soviet Mathematics Doklady, 1983, 27(2):372-376.
- [PK2012] Panagakis Y, Kotropoulos C. Automatic music tagging by low-rank representation. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2012, 497-500.
- [PGWXM2012] Peng Y, Ganesh A, Wright J, Xu W, Ma Y. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2233-2246.
- [QYZL2014] Jianjun Qian, Jian Yang, Fanlong Zhang, and Zhouchen Lin, Robust Low-Rank Regularized Regression for Face Recognition with Occlusion, CVPR 2014 Biometrics Workshop.
- [RL2013] Ren X, Lin Z. Linearized alternating direction method with adaptive penalty and warm starts for fast solving transform invariant low-rank textures. International J. Computer Vision, Vol. 104, pp. 1-14, 2013.
- [TFFWZ2013] Huachun Tan, Jianshuai Feng, Guangdong Feng, Wuhong Wang, and Yu-Jin Zhang, Traffic Volume Data Outlier Recovery via Tensor Model, Mathematical Problems in Engineering, Volume 2013, 2013.
- [VMS2015] Rene Vidal, Yi Ma, and S. S. Sastry, Generalized Principal Component Analysis, Springer, in press.
- [WSC2011] Joseph Wang, Venkatesh Saligrama, David Castanon, Structural Similarity and Distance in Learning, Annual Allerton Conf. Communication, Control and Computing 2011.
- [WL2010] Wei S, Lin Z. Analysis and improvement of low rank representation for subspace segmentation. arXiv:1107.1561, 2010.
- [WYZ2012] Zaiwen Wen, Wotao Yin, Yin Zhang: Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. Math. Program. Comput. 4(4): 333-361, 2012.
- [WGRM2009] Wright J, Ganesh A, Rao S, Ma Y. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. In: Proceedings of Advances in Neural Information Processing Systems (NIPS), 2009, 2080-2088.
- [WGSWM2010] Wu L, Ganesh A, Shi B, Matsushita Y, Wang Y, Ma Y. Robust photometric stereo via low-rank matrix completion and recovery. In: Proceedings of Asian Conference on Computer Vision (ACCV), 2010, vol. 3, 703-717.
- [YLLZ2014] Li Yang, Yang Lin, Zhouchen Lin, and Hongbin Zha, Low Rank Global Geometric Consistency for Partial-Duplicate Image Search, ICPR 2014.
- [YS2011] Yao-Liang Yu, Dale Schuurmans. Rank/Norm Regularization with Closed-Form Solutions: Application to Subspace Clustering. UAI 2011.
- [ZGLM2012] Zhang Z, Ganesh A, Liang X, Ma Y. TILT: Transform invariant low-rank textures. International Journal of Computer Vision, 2012, 99(1): 1-24.
- [ZGLA2012] Zhang T, Ghanem B, Liu S, Ahuja N. Low-rank sparse learning for robust visual tracking. In: Proceedings of European Conference on Computer Vision (ECCV), 2012, 470-484.
- [ZLSM2013] Xin Zhang, Zhouchen Lin, Fuchun Sun, and Yi Ma, Rectification of Optical Characters as Transform Invariant Low-rank Textures, ICDAR2013.
- [ZLSM2014] Xin Zhang, Zhouchen Lin, Fuchun Sun, and Yi Ma, Transform invariant text extraction, The Visual Computer, Vol. 30, No. 4, pp. 401-415, April 2014.
- [ZLZ2013] Hongyang Zhang, Zhouchen Lin, and Chao Zhang. A Counterexample for the Validity of Using Nuclear Norm as a Convex Surrogate of Rank. ECML/PKDD 2013.
- [ZLZC2015] Hongyang Zhang, Zhouchen Lin, Chao Zhang, and Edward Chang. Exact Recoverability of Robust PCA via Outlier Pursuit with Tight Recovery Bounds. AAAI 2015.
- [ZLZG2015] Hongyang Zhang, Zhouchen Lin, Chao Zhang, and Junbin Gao, Relation among Some Low Rank Subspace Recovery Models, submitted to Neural Computation.

- [ZLZG2014] Hongyang Zhang, Zhouchen Lin, Chao Zhang, and Junbin Gao. Robust Latent Low Rank Representation for Subspace Clustering. *Neurocomputing*, Volume 145, pp. 369-373, December 2014.
- [ZLM2011] Zhang Z, Liang X, Ma Y. Unwrapping low-rank textures on generalized cylindrical surfaces. In: *Proceedings of International Conference on Computer Vision (ICCV)*, 2011, 1347-1354.
- [ZMM2011] Zhang Z, Matsushita Y, Ma Y. Camera calibration with lens distortion from low-rank textures. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011: 2321-2328.
- [ZLWCM2010] Zhou Z, Li X, Wright J, Candès E J, Ma Y. Stable principal component pursuit. In: *International Symposium on Information Theory (ISIT)*, 2010, 1518 -1522.
- [ZYM2010] Zhu G, Yan S, Ma Y. Image tag refinement towards low-rank, content-tag prior and error sparsity. In: *Proceedings of ACM Multimedia*, 2010, 461-470.
- [ZGLMZY2012] Zhuang L, Gao H, Lin Z, Ma Y, Zhang X, Yu N. Non-negative low rank and sparse graph for semi-supervised learning. In: *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, 2328-2335.
- [ZL2011] Zuo W, Lin Z. A generalized accelerated proximal gradient approach for total-variation-based image restoration. *IEEE Transactions on Image Processing*, 2011, 20(10):2748-2759.