LaTeX Author Guidelines for ICCV Proceedings

Anonymous ICCV submission

Paper ID ****

Abstract

我们从*Weighted Orthogonal Procrustes Problem*出发，设计出一个*weighted Frobenious norm*作为*data term*，用*weighted sparse coding*作为*regularization term*。这个新的模型用*alternative updating*方式求解，可以证明其对于每个变量都有闭合解，并且收敛到一个*stationary point*。这个模型用在了真实去噪问题里，对之前咱提出的*guided*的方法在*60*张真实噪声图上的去噪效果更好，*PSNR*提升了*1.1dB*左右。

1. Introduction

The camera noise would be modeled as mixed Poisson and Gaussian noise, and the Poisson noise is usually very small and could be approximated by Gaussian. The noise in darker area of images would be amplified to have relatively larger variances while the noise in brighter area would be supressed to have relatively smaller variances. This makes the real noisy images to have local Gaussian noise in different area with different variances. Why the local noise within an image patch can be modeled as a Gaussian? We first introduce a simplified model including various noise sources (for each pixel):

$$\mathbf{P} = f((g_{cv}(\mathbf{C}+\mathbf{D})+\mathbf{N}_{reset})g_{out}+\mathbf{N}_{out})+\mathbf{Q} \quad (1)$$

$\mathbf{C}$ is the number of absorbed electrons (charges) transformed from Photons via Photon-diode which follows Poisson distribution. $\mathbf{D}$ is the number of absorbed electrons (charges) generated by Dark Current (thermal generation) which also follows Poisson distribution. Sum of Independent Poisson Random Variables is also Poisson. $\mathbf{N}_{reset}$ is the thermal noise generated by the readout circuitry (or reset noise related to reset voltage) and is currently admitted to be modeled as Gaussian distribution in the literature. $\mathbf{N}_{out}$ is the readout noise and also follows Gaussian distribution. Sum of Independent Gaussian Random Variables is also Gaussian $\mathbf{Q}$ is the quantization error (often uniformly distributed) happened during rounding to interger values and could be negligible compared to the readout noise. Pixels in a local region could share the same parameters for the noise emerged from the Poisson distribution since they are from the same point lighting source. Then the noise in these pixels has closed relationship with each other while the noise from distant pixels have weak or even no relationship. If the number of pixels is larger, the Poisson distribution can be more closedly approximated by a certain Gaussian distribution according to the famous Central Limit Theorem.

But one may doubt that if this is the fact, then the overall noise model of any real noisy image could be approximated by Gaussian distribution. That is, the Gaussian distribution of noise profile should be a global property instead of local one.

Traditional Frobenius norm has been successfully accounted for Gaussian noise from the Bayesian perspective. However, the Frobenius norm is not robust to outliers or fails to account for non-Gaussian noise, which are commonly existed in realistic applications. For example, the real-world noisy images are mainly degraded by mixed Poisson and Gaussian noise, and the traditional Frobenius norm is not reasonable or effective to real noisy image denoising problem. In order to deal with outliers and non-

Gaussian noise, the ℓ_1 norm of matrix or the robust function [1] could be employed to replace the traditional Frobenius norm. However, this will make the overal l problem more complex since the problem with ℓ_1 or robust function are non-smooth and could hardly have closed-form solutions. For example, the bilinear factorization problem with weights [2] is NP-hard and does not have a closed-form solution in general.

The extension from unweighted norms to weighted ones can ba traced back to the work of [3] on factor analysis. Weights are frequently used in many computer vision and machine learning applications. For example, the weights of indicator matrix is frequently used in image restoration [] or matrix factorization with missing entries [2].

Two different weighted Frobenius norms are defined by Higham in his paper [4]. The (1.3) of [4] is the Hadamard weighted Frobenius norm while the (1.3) of [4] is the most commonly used weighted Frobenius norm in numerical mathematics, where $\mathbf{W}$ is the symmetric positive definite matrix for the weighted Frobenius norm. The weighted Frobenius norm models are also more robust to outliers than unweighted one since an outlier belong to the local patch can be dealed by a Gaussian with large variance [5].

Currently, the weighted Frobenius norm is applied in the weighted low rank approximations (WLRA) problem [2] and the rubust sparse coding problem for image processing [6] and pattern recognition [5]. Srebro and Jaakkola [2] proposed the weighted low-rank approximation model which employed the EM algorithm or conjugate gradient descent algorithm to solve the LRMF problem. However, the weighted LRMF problem may have local minimum which is not the global when compared to the unweighted canse [2]. Meng *et al.* [5] solved the WLRA via Expectation-Maximization (EM) algorithm by viewing WLRA problem as a maximum-likelihood estimation (MLE) problem with missing entries.

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{U}\mathbf{V}^\top)\|_F \tag{2}$$

If $p = 1$, this problem is a NP-hard problem and can have several solutions (local minimum?) [7], this may be attributed to the lack of convexity of the norm in terms of matrices of $\mathbf{U}$ and $\mathbf{V}$. The parametrs of the weighting matrix $\mathbf{W}$ are obtained by the estimated parameters of the MoG model during the alternative updating steps in Expectation-Maximization (EM) algorithm under the maximum likelihood estimation (MLE) framework. When the weighting matrix is predefined, existing algorithms, such as the Alternated Least Squares (ALS) [8], WLRA [2], and the Damped Newton algorithms [9] can be employed to solve this weighted L2 LRMF problem.

这个模型应用在Low-Rank Matrix Factorization with Missing Entries这个问题里，$\mathbf{W}$指矩阵的元素是否是missing entry，其元素只有0或1。这几篇文章中，都用Mixture of Gaussians (MoG)模拟噪声，用EM算法估计噪声分布，$\mathbf{W}_{ij}$是噪声属于所有Gaussian的期望与高斯噪声方差的比值的和，即$\mathbf{W}_{ij} = \sum_{k=1}^{K} \frac{\gamma_{ijk}}{2\sigma_k^2}$。这个加权方式不适合用于对真实噪声加权。原因是1，这个权重矩阵变量太多，迭代的计算量很大。2，每个元素对应于一个像素值，但是很难仅仅通过一个像素值就估计出这个像素值里噪声的水平。3，真实噪声里，可能某个像素的噪声是outlier，而EM算法对outlier不鲁棒，导致无法很好地去除这个像素的噪声。4，EM算法估计过程计算量很大，不适合用于对真实噪声加权。5，对单个像素加一个权重忽略了图像的结构信息，这个加权方式适用于随机程度比较高的问题，比如矩阵分解，但是对于图像修复或者分类就不太适用。

杨猛的模型[10]主体是

$$\min_{\mathbf{c}} \|\mathbf{W}^{\frac{1}{2}}(\mathbf{y} - \mathbf{D}\mathbf{c})\|_2^2 \quad \text{s.t.} \quad \|\mathbf{c}\|_1 < \sigma \tag{3}$$

用权重矩阵左乘redidual 向量，其物理含义是对residual向量的每个variable加权，这是因为这个模型主要用在人脸识别上，当人脸有遮挡时，这个权重可以把遮挡部分的像素值的权重变小，使得模型对于有遮挡的人脸识别依然有效。这个加权方式不适合用于对真实噪声加权。原因是1，在真实噪声去噪问题里，如果把一些相似块堆起来组成一个矩阵，每一行是没有任何空间位置关系的点，要求这些点上的噪声符合高斯分布是没有依据的（相当于在图像中放一个网格，要求网格上的所有交叉点所对应的像素上的噪声符合高斯分布是不合理的)。

我们把Frobenious Norm里的residual矩阵（注意是矩阵，而不是向量）右乘对角的权重矩阵$\mathbf{W}$，相当于

假设真实噪声图的每个局部的噪声都可以近似于一个高斯，不同的局部的噪声水平不同，对应的权重不同。我们还学习得到一个正交字典，正交字典的好处后面会提到。因为正交字典里的每一列的atom是不同重要程度的，这个重要程度可以通过奇异值矩阵$\mathbf{S}$对角线上的对应奇异值的大小来决定，因此，我们还在residual矩阵里正则项系数$\mathbf{C}$的左边乘上一个$\mathbf{S}$来对正则项的每一行加上一个权重。而且实验中，我们发现，加权矩阵的引入使得模型可以自适应于不同的噪声情况，从而使得我们的模型收敛速度更快。

2. Weighted Orthonormal Dictionary Learning

现有一个图像块，假设是8x8($d = 64$)的维度，拉成列向量$\mathbf{y} \in \mathbb{R}^d$，那么我们在其周围找$m$个相似块，得到相似块矩阵$\mathbf{Y} \in \mathbb{R}^{d\times m}$。我们希望利用稀疏表达这个工具。很自然，我想到了K-SVD这个模型，给定数据矩阵$\mathbf{Y}$,KSVD不断迭代更新，从而得到最适合表达数据的字典和稀疏。

与KSVD不同的是，我们要求这个新模型里的字典是正交的。原因是：1，相似块在一个低维空间里，所以正交基就足够了，不需要用过完备基；2，我们模型里的正交字典正好有闭合解，并且结合加权sparse coding得到的整个模型是有收敛性保证的。从而，我们想到的模型主体是这样的：

$$\min_{\mathbf{D},\mathbf{C}} \frac{1}{2}\|(\mathbf{Y}-\mathbf{DC})\mathbf{W}\|_F^2 + \lambda\|\mathbf{S}^{-1}\mathbf{C}\|_1 \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D} = \mathbf{I}. \tag{4}$$

其中$\mathbf{S}$是对角正定矩阵(表示sparse coding每一行权值不同)。定义 $\mathbf{C}^* = \mathbf{S}^{-1}\mathbf{C}$, 有$\mathbf{C} = \mathbf{S}\mathbf{C}^*$ and

$$\min_{\mathbf{D},\mathbf{C}^*} \frac{1}{2}\|(\mathbf{Y}-\mathbf{DSC}^*)\mathbf{W}\|_F^2 + \lambda\|\mathbf{C}^*\|_1 \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D} = \mathbf{I}. \tag{5}$$

权重矩阵$\mathbf{W}$的设计应该遵循如下原理：噪声高，权重就小；噪声低，权重就大。这个原则是从求解系数矩阵$\mathbf{C}$时的贝叶斯法则得来的。
一方面，求解系数矩阵$\mathbf{C}$需要每列每列单独求解：

$$\hat{\mathbf{c}}_i^* = \arg\min_{\mathbf{c}_i^*} \frac{1}{2}\|(\mathbf{y}_i - \mathbf{DSc}_i^*)\mathbf{W}_{ii}\|_2^2 + \lambda\|\mathbf{c}_i^*\|_1. \tag{6}$$

另一方面，由MAP框架，有

$$\hat{\mathbf{c}}_i^* = \arg\max_{\mathbf{c}_i^*} \ln P(\mathbf{c}_i^*|\mathbf{y}_i). \tag{7}$$

由贝叶斯法则，等价于

$$\hat{\mathbf{c}}_i^* = \arg\max_{\mathbf{c}_i^*} \{\ln P(\mathbf{y}_i|\mathbf{c}_i^*) + \ln P(\mathbf{c}_i^*)\}. \tag{8}$$

其中拟合residual noise的数据项是

$$P(\mathbf{y}_i|\mathbf{c}_i^*) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp(-\frac{1}{2\sigma_i^2}\|\mathbf{y}_i - \mathbf{DSc}_i^*\|_2^2). \tag{9}$$

系数正则项是Laplace distribution with location $\mathbf{0}$ and scale $\lambda > 0$:

$$P(\mathbf{c}_i^*) = \prod_{j=1}^{d}\frac{\lambda}{2}\exp(-\lambda|\mathbf{c}_{ij}^*|). \tag{10}$$

从而得到

$$\hat{\mathbf{c}}_i^* = \arg\min_{\mathbf{c}_i^*} \frac{1}{2}\|(\mathbf{y}_i - \mathbf{DSc}_i^*)\frac{1}{\sigma_i}\|_2^2 + \lambda\|\mathbf{c}_i^*\|_1. \tag{11}$$

对比这两个方面得到的系数矩阵，有

$$\mathbf{W}_{ii} = \frac{1}{\sigma_i} \tag{12}$$

也就是说，权重应该是噪声水平的倒数：噪声高，权重就小；噪声低，权重就大。

3. 初始化

- $\mathbf{W}$上的对角元素是相似块矩阵$\mathbf{Y}$每一列的初始噪声水平的倒数。注意$\mathbf{Y}$已经去掉DC components，列均值为$\mathbf{0}$。因为$\mathbf{Y}$的每一列不全是噪声，还有信号部分，所以初始噪声水平近似于把$\mathbf{Y}$的每列乘上一个系数λ_{ls}，得到权重矩阵$\mathbf{W}_{ii} = \frac{1}{\lambda_1\sqrt{var(\mathbf{y}_i)}}$。因为每一列中信号所占比例很难区别对待，所以乘以同一个系数。之前的做法是初始化为恒等矩阵$\mathbf{I}$(原因是一开始把所有块平等对待，视为受到相同的高斯噪声的影响)，这对于真实噪声是不合理的；或者初始化为一个对角矩阵，对角线上的每个元素对应于$\mathbf{Y}$的每一列（图像块）上的噪声水平$\sigma_i = \lambda_1\sqrt{var(\mathbf{y}_i)}$，这个方法弄反了，应该是噪声水平的倒数。
- 正交字典$\mathbf{D}^{(0)} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}$和稀疏系数$\mathbf{C}$行方向上的权值$\mathbf{S} = \mathbf{\Sigma}^{(0)}$由$(\mathbf{Y}\mathbf{W}^{(0)})(\mathbf{Y}\mathbf{W}^{(0)})^\top = \mathbf{U}\mathbf{\Sigma}^{(0)}\mathbf{U}^\top$做SVD分解得到，$\mathbf{C}$每一行的权值不同，意味着正交字典$\mathbf{D}$的每一列对$\mathbf{C}$的重要性不同；

4. 反复迭代求解D,C直到收敛,更新W

4.1. 反复迭代求解D,C直到收敛

$k = 0, 1, 2, ...$:

a. update $\mathbf{C}$

$$\min_{\mathbf{C}^*} \frac{1}{2}\|(\mathbf{Y} - \mathbf{D}^{(k)}\mathbf{S}\mathbf{C}^*)\mathbf{W}\|_F^2 + \lambda\|\mathbf{C}^*\|_1. \quad (13)$$

有闭合解，每一列单独求解：

$$(\hat{\mathbf{c}}_i^*)^{(k+1)} = \arg\min_{\mathbf{c}_i} \frac{1}{2}\|(\mathbf{y}_i - \mathbf{D}^{(k)}\mathbf{S}\mathbf{c}_i^*)\mathbf{W}_{ii}\|_2^2 + \lambda\|\mathbf{c}_i^*\|_1. \quad (14)$$

闭合解为：

$$(\hat{\mathbf{c}}_i^*)^{(k+1)} = \mathrm{sgn}(\mathbf{S}\mathbf{D}^\top\mathbf{y}) \odot \max(|\mathbf{S}\mathbf{D}^\top\mathbf{y}| - \frac{\lambda}{(\mathbf{W}_{ii})^2}, 0), \quad (15)$$

别忘了$\mathbf{C}^{(k+1)} = \mathbf{S}(\mathbf{C}^*)^{(k+1)}$!

b. update $\mathbf{D}$

$$\min_{\mathbf{D}} \frac{1}{2}\|(\mathbf{Y} - \mathbf{D}\mathbf{C}^{(k+1)})\mathbf{W}\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D} = \mathbf{I}. \quad (16)$$

等价于

$$\min_{\mathbf{D}} \|(\mathbf{Y}\mathbf{W}) - \mathbf{D}(\mathbf{C}^{(k+1)}\mathbf{W})\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D} = \mathbf{I}, \quad (17)$$

闭合解为：$\hat{\mathbf{D}}^{(k+1)} = \mathbf{V}\mathbf{U}^\top, \mathbf{C}\mathbf{W}(\mathbf{Y}\mathbf{W})^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$.

需要证明收敛性，用ADMM方法证明，并给一个收敛的图：

$$\min_{\mathbf{D},\mathbf{C}^*} \|(\mathbf{D}^\top\mathbf{Y} - \mathbf{S}\mathbf{C}^*)\mathbf{W}\|_F^2 + \lambda\|\mathbf{C}^*\|_1 \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D} = \mathbf{I}. \quad (18)$$

实验中，我们发现，加权矩阵的引入使得模型可以自适应于不同的噪声情况，从而使得我们的模型收敛速度更快。如果不加权重矩阵$\mathbf{W}$， $\mathbf{S}$，那么收敛速度会慢很多。

Since the objective function of (4) is monotonically non-increasing and has the lower bound (0), it is convergent according to the famous Monotone Convergence Theorem [11] http://math.stackexchange.com/questions/789521/non-increasing-monotone-sequence-convergence-proof. It could be reasonably terminated when the decreasing rate is smaller than a preset threshold or the number of alternating updating steps of the optimization problem reaches the maximum iteration number.

4.2. WOPP收敛后，更新两个权重矩阵$\mathbf{W}$， $\mathbf{S}$

c. update $\mathbf{W}$

稀疏系数$\mathbf{C}$列方向上的权重矩阵$\mathbf{W}$是对角矩阵，只需要更新迭代对角元即可。有两种迭代方式：

Noise Level Case：

$$\mathbf{W}_{ii}^{new} = \frac{\lambda_2}{\sigma_i - \|\mathbf{y}_i - \mathbf{D}^{(k+1)}\hat{\mathbf{c}}_i^{(k+1)}\|_2} \quad (19)$$

λ_2是参数；

*RBF Case在这里是不合理的设计方式，可作为备用：

$$\mathbf{W}_{ii}^{new} = \exp(\lambda_2\|\mathbf{y}_i - \mathbf{D}^{(k+1)}\hat{\mathbf{c}}_i^{(k+1)}\|_2) \quad (20)$$

d. update $\mathbf{S}$

稀疏系数$\mathbf{C}$行方向上的权重矩阵$\mathbf{S}$是对角矩阵：$\mathbf{S} = \mathbf{\Sigma}$ or $\mathbf{S} = \sqrt{\mathbf{\Sigma}}$由$(\mathbf{Y}\mathbf{W}^{new})(\mathbf{Y}\mathbf{W}^{new})^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top$做SVD分解得到。

References

[1] Peter J Huber. Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics*, pages 799–821, 1973. 2

[2] Nathan Srebro and Tommi Jaakkola. Weighted low-rank approximations. In *ICML*, volume 3, pages 720–727, 2003. 2

[3] Gale Young. Maximum likelihood estimation and factor analysis. *Psychometrika*, 6(1):49–53, 1941. 2

[4] Nicholas J Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002. 2

[5] Deyu Meng and Fernando De La Torre. Robust matrix factorization with unknown noise. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1337–1344, 2013. 2

[6] Jielin Jiang, Lei Zhang, and Jian Yang. Mixed noise removal by weighted encoding with sparse nonlocal regularization. *IEEE transactions on image processing*, 23(6):2651–2662, 2014. 2

[7] Nicolas Gillis and François Glineur. Low-rank matrix approximation with weights or missing data is np-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1149–1165, 2011. 2

[8] Fernando De La Torre and Michael J Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117–142, 2003. 2

[9] Aeron M Buchanan and Andrew W Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 316–322. IEEE, 2005. 2

[10] Meng Yang, Lei Zhang, Jian Yang, and David Zhang. Robust sparse coding for face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 625–632. IEEE, 2011. 2

[11] Elias M Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2009. 4