

基于矩阵分解的荧光显微序列线粒体检测

贾晓萌, 聂永明, 都思丹

(南京大学电子科学与工程学院, 江苏南京 210023)

摘要:利用图像处理技术,检测荧光显微序列中的活性线粒体是生物医学领域重要的研究手段之一.受到荧光显微镜成像技术的限制,序列中每帧图像均包含细胞质阴影和荧光标记的线粒体,具有很低的信噪比,难以满足一般粒子检测算法的要求.为了精确检测活细胞中的线粒体,提出一种基于矩阵分解的荧光显微序列线粒体检测算法,并利用增广拉格朗日乘子法,快速准确地实现该算法,将线粒体从细胞质阴影中有效分离出来,实现线粒体的精确检测.实验结果表明,此方法为活细胞中线粒体的精确检测提供了快速、高效的分析工具.

关键词:线粒体;矩阵分解;优化问题;增广拉格朗日乘子法

中图分类号:R319 **文献标识码:**A **doi:**10.3969/j.issn.0253-2778.2014.10.007

引用格式: Jia Xiaomeng, Nie Yongming, Du Sidan. A detection algorithm based on matrix factorization for live mitochondria in fluorescent microscopic images[J]. Journal of University of Science and Technology of China, 2014, 44(10): 839-843.

贾晓萌, 聂永明, 都思丹. 基于矩阵分解的荧光显微序列线粒体检测[J]. 中国科学技术大学学报, 2014, 44(10): 839-843.

A detection algorithm based on matrix factorization for live mitochondria in fluorescent microscopic images

JIA Xiaomeng, NIE Yongming, DU Sidan

(School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China)

Abstract: Detection of mitochondria in fluorescent microscopic images is one of the most important methods in studies concerning apoptosis and the nature of life phenomena in the area of biomedical image processing. Limited by fluorescence microscopy, fluorescent microscopic images contain two parts which are the shadow of cytoplasm and live mitochondria, and the signal-to-noise ratio (SNR) of live mitochondria time sequence images is low, which cannot meet the requirements of general particle algorithm. A new detection algorithm was proposed for live mitochondria in fluorescent microscopic images. To realize this method rapidly, augmented Lagrange multiplier algorithm was used. Mitochondria was be separated from the cytoplasm and accurately detected in fluorescent microscopic images. Therefore, the proposed algorithm provides an efficient and accurate tool to detect mitochondria in live cell.

Key words: mitochondria; matrix factorization; optimization problem; augmented Lagrange multiplier algorithm

收稿日期:2014-03-18;修回日期:2014-05-16

基金项目:国家自然科学基金(61271231,61300157,61201425),江苏省自然科学基金(BK2011337)资助.

作者简介:贾晓萌,女,1988年生,硕士生.研究方向:图像处理. E-mail: jxm88922@163.com

通讯作者:都思丹,博士/教授. E-mail: coff128@nju.edu.cn.

0 引言

自动检测生物荧光显微图像序列中的活性线粒体,为科学家进一步研究生物活细胞中线粒体的运动情况,了解细胞的凋亡过程,生命现象的本质等方面提供了基础的研究保障.2006年,Conrad^[1]的研究表明了线粒体迁移对细胞调控机制的影响,引起了广泛的关注.目前,在生物荧光显微图像序列中检测活性线粒体,是生物图像处理领域的研究热点.

受显微镜光学成像技术的限制和荧光蛋白光漂白作用等因素的影响,使获取的生物光学显微图像的信噪比无法达到普通粒子识别算法所需的条件.在光学显微图像中,用绿色荧光蛋白(green fluorescent protein, GFP)标记的线粒体,通常在非均匀背景中呈现为附着于轴突的密集小尺寸粒子,这些粒子没有明显的边界并且有不均匀的灰度分布^[2],所以精确检测出荧光显微图像中线粒体的位置的挑战,主要有以下两个方面:

(I) 缺乏可靠的检测特征:在一副图像中,对于感兴趣的检测目标,设计检测算法的关键是选取易于区分的目标特征,如检测目标的颜色、强度、边缘等.由于线粒体在显微图像中呈现为没有明显边界的密集微粒,缺乏可靠的检测特征,因此想要进行精确的检测是非常困难的.

(II) 背景干扰的影响:由于荧光显微成像技术的发展并不完善,通常获取的生物图像的信噪比很低,且受到细胞质阴影的干扰,背景复杂度很高,难以达到传统粒子检测算法的要求.

在活细胞成像中,自动化的粒子检测是一项非常具有挑战性的任务.目前,已经提出很多种方法用于完成这项任务.Olivo-Marin^[3]提出了多尺度小波变换的方法,使用 à trous 小波变换将目标粒子从原始图像中检测出来,这种算法在图像信噪比较高时可以准确地检测图像中粒子的位置分布,但当采集到的图像质量变差时,则难以获得准确的检测结果.Jiang^[4]提出一种机器学习方法,这种方法基于 Haar 特征检测,可以在图像中检测出一部分快速运动的目标粒子.

Baraniuk^[5]和 Donoho^[6]最近提出的压缩感知(compressed sensing)理论在信号处理和图像处理领域得到了广泛的应用.这一理论突破了经典的 Nyquist-Shannon 采样定理,充分利用现实中大多数信号是稀疏的这一特性,在保证准确重构出原始

信号的前提下,大大降低了采样所需的采样率.在压缩感知理论的基础上,进一步提出矩阵分解理论,这个理论可以将一个高维数的信号准确地分解成低秩(low-rank)成分和稀疏(sparse)成分,为信号的进一步处理做准备.

本文提出一种在荧光显微图像序列中,对于活性线粒体检测非常有效的自动检测算法.这种算法结合压缩感知领域的矩阵恢复算法,将一个图像序列中的每帧作为数字矩阵的一列.由于图像序列中帧与帧之间具有很大的相似性,则该矩阵具有较低的秩.根据这个特性,矩阵恢复算法可以将每帧中的相似部分和特有部分分别当作数据矩阵的低秩成分和稀疏成分进行分解.每帧图像中灰度分布不均匀的细胞质背景可被看作是低秩成分,而我们感兴趣的线粒体则被看作是稀疏成分.低秩成分和稀疏成分的分解,实现了线粒体和细胞质阴影的分离,显微图像的信噪比和背景复杂度都得到了明显的改善,满足一般粒子检测算法的要求,使得运动的线粒体可以被精确地检测出来.

1 基于矩阵分解线粒体检测模型

本文使用的数据是果蝇神经元轴突内的线粒体荧光显微序列.序列中每帧图像上的神经元轴突均包含荧光标定后的线粒体.由于在细胞荧光显微成像时,轴突细胞中的细胞质会呈现阴影,对精确检测线粒体的位置造成了很大的干扰.根据观察图像序列可以发现,细胞质在每帧图像中呈现的阴影具有很大的相似性.并且根据 Pilling^[7]的研究结果,在果蝇运动神经轴突中,有 57% 的线粒体是不运动的,仅有 43% 的线粒体是运动的.

假设每帧图像 D_i 可以分解为神经轴突的细胞质阴影 A 和荧光标定线粒体 E 两部分,则检测模型可以表示为:

$$D = A + E \quad (1)$$

如果将每帧图像 D_i 中的像素排列成一列作为一个数字矩阵 D 的列,则该矩阵的秩就是 1,并且分解之后的 E 中只有很少一部分元素是非零的,即 E 是稀疏的.为了精确分解式(1),根据已知背景,我们设定目标函数为 $\text{rank}(A) + \lambda \|E\|_0$,由此就可以将矩阵分解问题转化为最优化问题.Candès^[8]指出这个优化问题称为 robust PCA(RPCA),可以确保将矩阵的低秩成分 A 和稀疏成分 E 分解开,完成剔除细胞质阴影 A ,获得只含有线粒体 E 的图像的目的.

2 模型求解算法

矩阵恢复 (matrix recovery, MR)^[8-9] 又称为 RPCA 或者稀疏与低秩矩阵分解, 是指当矩阵的某些元素被严重“污染”后, 可以自动识别出被“污染”的元素 (实际上被“污染”的元素即是我们希望检测出的线粒体), 并恢复出原矩阵的方法. 其中, 假定原矩阵是低秩的, 而被严重“污染”的元素是稀疏的, 于是矩阵恢复可用如下优化问题来表达:

$$\begin{aligned} \min \quad & \text{rank}(\mathbf{X}) + \lambda \|\mathbf{E}\|_0 \\ \text{subject to} \quad & \mathbf{A} + \mathbf{E} = \mathbf{D} \end{aligned} \quad (2)$$

式中, $\|\cdot\|_0$ 表示一个矩阵的 l_0 范数, λ 表明噪声所占的权重. 这个式子在理论和实践中, 计算复杂度都是指数级的, 因此可以通过松弛技巧, 用 l_1 范数代替 l_0 范数. 即, 用核范数近似矩阵的秩, 并用矩阵的 l_1 范数来近似矩阵的零范数, 则原问题转化为:

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{subject to} \quad & \mathbf{A} + \mathbf{E} = \mathbf{D} \end{aligned} \quad (3)$$

式中, $\|\cdot\|_*$ 和 $\|\cdot\|_1$ 分别表示一个矩阵的核范数和 l_1 范数, 定义为 $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$, $\sigma_i(\mathbf{X})$ 表示矩阵的第 i 个奇异值, $\|\mathbf{X}\|_1 = \sum_{i,j} \mathbf{E}_{ij}$, 表示矩阵 \mathbf{E} 中所有元素相加. 这是一个凸优化问题, 但是由于问题所涉及的维数非常大, 所以如何快速求解该问题是非常关键的. 这里我们运用增广拉格朗日法求解式(3).

矩阵恢复问题的增广拉格朗日函数可以表示为:

$$\begin{aligned} L(\mathbf{A}, \mathbf{E}, \mathbf{Y}, \mu) = & \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + \\ & \langle \mathbf{Y}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\mu}{2} \|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \end{aligned} \quad (4)$$

式中, $\|\cdot\|_F$ 表示 Frobenius 范数. 要将式(4)进行最小化, 可以利用交替更新的方式, 即先固定 \mathbf{E} 和 \mathbf{Y} , 求一个使得 $L(\cdot)$ 最小化的 \mathbf{A} , 然后再固定 \mathbf{A} 和 \mathbf{Y} , 求一个使 $L(\cdot)$ 最小的 \mathbf{E} , 按照这样的方式进行更新, 就可以收敛得到该问题的最优解.

更新 \mathbf{A} 时,

$$\begin{aligned} \text{argmin}_{\mathbf{A}} \quad & \|\mathbf{A}\|_* + \frac{\mu}{2} \|\mathbf{D} - \mathbf{A} - \mathbf{E} + \mu^{-1}\mathbf{Y}\|_F^2 = \\ & \mathbf{D}_{\mu^{-1}}^{-1}(\mathbf{D} - \mathbf{E} + \mu^{-1}\mathbf{Y}) \end{aligned} \quad (5)$$

更新 \mathbf{E} 时,

$$\text{argmin}_{\mathbf{E}} \quad \|\mathbf{E}\|_1 + \frac{\mu}{2} \|\mathbf{D} - \mathbf{A} - \mathbf{E} + \mu^{-1}\mathbf{Y}\|_F^2 =$$

$$\mathbf{S}_{\frac{\lambda}{\mu}}(\mathbf{D} - \mathbf{E} + \mu^{-1}\mathbf{Y}) \quad (6)$$

根据式(5)和(6)进行交替更新, 直到原问题收敛. 实际上, 并不需要在每一步中都求解子问题的精确解, 只需更新 \mathbf{A} 与 \mathbf{E} 各一次得到子问题的一个近似解, 已经足以使算法最终收敛得到原问题的最优解. 这样, 可以得到一个收敛速度更快的解法, 称为增广拉格朗日乘法 (augmented Lagrange multiplier). 其中, $\mathbf{D}(\cdot)$ 代表奇异值分解 (singular value decomposition), 在算法中用 $\text{svd}(\cdot)$ 表示, $\mathbf{S}(\cdot)$ 是软阈值标量 (soft threshold) 算子, 定义为:

$$\mathbf{S}_{\epsilon}(x) = \begin{cases} \text{sign}(x)(|x| - \epsilon), & |x| > \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

增广拉格朗日乘法可以详细表示如下:

算法 2.1 增广拉格朗日乘法

初始值: $\mathbf{Y}_0, \mathbf{E}_0 = 0, \mu_0, \mathbf{k} = 0, \lambda = 1600$

While 直到收敛

$(\mathbf{U}, \mathbf{S}, \mathbf{V}) = \text{svd}(\mathbf{D} - \mathbf{E}_{\mathbf{k}} - \mu_{\mathbf{k}}^{-1}\mathbf{Y}_{\mathbf{k}})$

$\mathbf{A}_{\mathbf{k}+1} = \mathbf{U}\mathbf{S}_{\mu_{\mathbf{k}}}^{-1}[\mathbf{S}]\mathbf{V}^T$

$\mathbf{E}_{\mathbf{k}+1} = \mathbf{S}_{\mu_{\mathbf{k}}}^{-1}[\mathbf{D} - \mathbf{E}_{\mathbf{k}} + \mu_{\mathbf{k}}^{-1}\mathbf{Y}_{\mathbf{k}}]$

$\mathbf{Y}_{\mathbf{k}+1} = \mathbf{Y}_{\mathbf{k}} + \mu_{\mathbf{k}}(\mathbf{D} - \mathbf{A}_{\mathbf{k}+1} - \mathbf{E}_{\mathbf{k}+1})$

$\mu_{\mathbf{k}} = \rho\mu_{\mathbf{k}-1}$

$\mathbf{k} = \mathbf{k} + 1$

end

3 实验

3.1 图像数据采集

本文分析的图像样本来自三期果蝇幼虫, 在缓冲液中迅速解剖暴露出神经束, 采用 Gal4/UAS 系统特异性在果蝇运动神经元中用定位绿色荧光蛋白 (mito-GFP) 表达线粒体, 常温下使用 Leica DM5000 宽场荧光显微镜拍摄所得. 该序列共有 60 帧, 每帧图像的分辨率为 997×801 像素. 在处理数据时, 将整个序列的 60 帧图像采集到一个数据矩阵中, 将每帧图像按照顺序作为数据矩阵中的一列, 这样就得到一个 $\mathbf{D} \in \mathbf{R}^{60 \times 798597}$ 的数据矩阵, 该矩阵理论上满足矩阵分解理论的条件假设, 即秩为 1.

3.2 图像数据分析

采用本文方法对序列中的 60 帧荧光显微图像进行处理, 从处理后的图像中随机选取 3 帧图像, 结果如图 1 所示. 图 1(a) 是原始的荧光显微图像; 图 1(b) 和图 1(c) 分别是原始图像中分解出的低秩成分和稀疏成分. 观察图 1 可知, 图 1(b) 中的低秩成分是对线粒体检测造成干扰的细胞质阴影, 阴影中

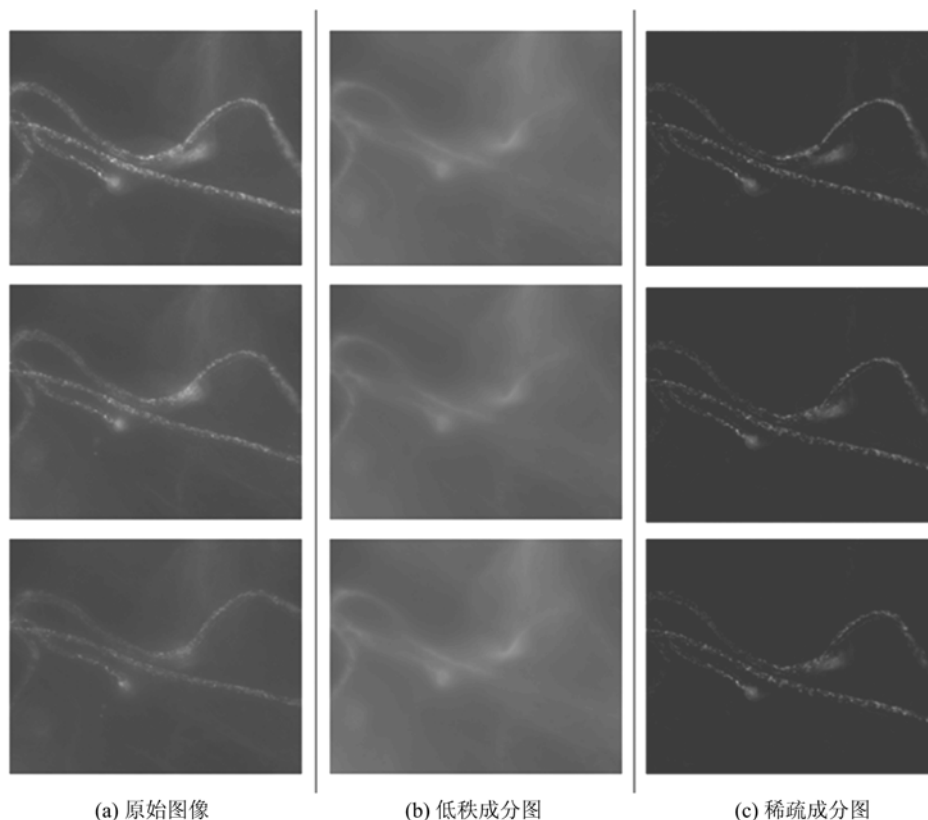


图 1 矩阵分解方法处理效果图

Fig. 1 Results of matrix factorization method

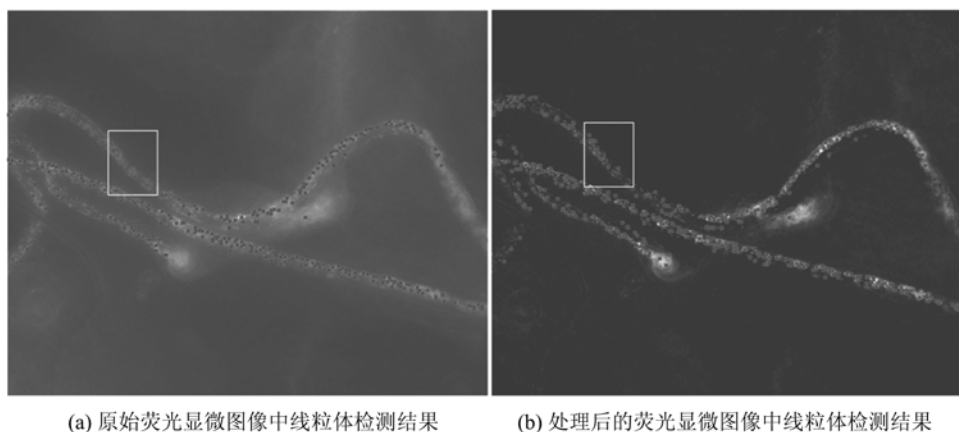


图 2 线粒体粒子检测结果对比图

Fig. 2 The comparison of mitochondria particle detection results

不包含任何线粒体;而在图 1(c)中仅包含我们感兴趣的、希望被检测出来的所有线粒体. 相对于图 1(a),图 1(c)中的细胞质阴影已基本被去除. 这样就实现了将线粒体从细胞质阴影中分离的目的,获得信噪比显著提高、细胞质干扰剔除干净、背景复杂度大幅度降低的显微图像. 处理后的显微图像的质量得到显著改善,在其中进行线粒体检测即可以

获得准确的检测结果. 在算法中平衡因子 λ 的选择非常重要,它决定了输出结果中稀疏成分和低秩成分所占的比例. 实验中各参数的选取如表 1 所示.

本文利用各向同性非抽样小波(IUWT)检测算法^[3]分别作用于原始显微图像和用矩阵分解方法处理后的显微图像中,检测结果如图 2 所示. 由于该算法在图像信噪比较高时可以获得准确的检测结果,

但随着图像的信噪比降低和图像背景复杂度的提高,其检测性能会随之快速下降,因此对比图 2(a)和图 2(b)中白色方框标出的区域,可以很明显地观察到,在图 2(a)中,由于轴突中细胞质阴影的干扰,白色方框区域内只有极少部分的线粒体可以被检测出来,大部分的线粒体都无法得到正确的检测;图 2(b)中,在完全去除干扰的显微图像中,白色方框区域内被检测出的线粒体数目明显增多,并且检测的线粒体位置也更准确.相比于直接在原始图像中检测线粒体,在矩阵分解方法处理后的图像中检测线粒体,可以得到更为准确的结果.

表 1 增广拉格朗日乘子算法参数值

Tab. 1 Parameters of augmented Lagrange multiplier algorithm

参 数	参数值
平衡因子 λ	1 600
加速因子 ρ	0.7
初始值 μ_0	2

4 结论

本文利用基于增广拉格朗日乘子法的矩阵恢复算法,将处于低信噪比荧光显微图像中的线粒体从细胞质阴影中分离出来,显著地提高了整个显微图像的图像质量,然后利用普通的粒子检测算法作用于处理后的图像中,即可快速且准确地检测出线粒体详细的位置分布情况,为进一步研究活细胞线粒体动力学特征提供了准确、高效的分析工具.

目前,在生物医学领域,自动跟踪荧光显微图像序列中活性线粒体的运动轨迹也是一个研究热点和难点,而自动检测活细胞中的线粒体是自动跟踪线粒体的基础,因此在自动检测线粒体的基础上完成对运动线粒体的自动跟踪是我们下一步的目标.

参考文献 (References)

- [1] Conradt B. Cell biology: Mitochondria shape up[J]. Nature, 2006, 443(7112): 646-647.
- [2] Xu J M, Li Y, Du S D, et al. A tracking algorithm for live mitochondria in fluorescent microscopy images[J]. Journal of Biomedical Engineering, 2012, 29(2): 352-358.
徐俊梅, 李杨, 都思丹, 等. 荧光显微图像序列中活性线粒体的跟踪算法[J]. 生物医学工程学杂志, 2012, 29(2): 352-358.
- [3] Olivo-Marin J C. Extraction of spots in biological images using multi-scale products [J]. Pattern Recognition, 2002, 35(9): 1 989-1 996.
- [4] Jiang S, Zhou X B, Kirchhausen T, et al. Detection of molecular particles in live cells via machine learning [J]. Cytometry A, 2007, 71(8): 563-575.
- [5] Baraniuk R. Compressive sensing [J]. IEEE signal processing magazine, 2007, 24(4): 118-121.
- [6] Donoho D L. Compressed sensing [J]. IEEE Transactions on Information Theory, 2006, 52(4): 1 289-1 306.
- [7] Pilling A D, Horiuchi D, Lively C M, et al. Kinesin-1 and Dynein are the primary motors for fast transport of mitochondria in Drosophila motor axons[J]. Molecular Biology of the Cell, 2006, 17(4): 2 057-2 068.
- [8] Candès E J, Li X D, Ma Y, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58(3): 1-39.
- [9] Lin Z C, Chen M M, Ma Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices [EB/OL]. http://www.researchgate.net/publication/46587364_The_Augmented_Lagrange_Multiplier_Method_for_Exact_Recovery_of_Corrupted_Low-Rank_Matrices.