Robust and Efficient Alignment of Calcium Imaging Data through Simultaneous Low Rank and Sparse Decomposition

Junmo Cho[1]*, Seungjae Han[1]*, Eun-Seo Cho[1], Kijung Shin[1,2], Young-Gyu Yoon[1,3]

[1]School of Electrical Engineering, KAIST
[2]Kim Jaechul Graduate School of AI, KAIST
[3]KAIST Institute for Health Science and Technology
Daejeon, Republic of Korea

{junmokane, jay0118, eunseo.cho, kijungs, ygyoon}@kaist.ac.kr

Abstract

Accurate alignment of calcium imaging data, which is critical for the extraction of neuronal activity signals, is often hindered by the image noise and the neuronal activity itself. To address the problem, we propose an algorithm named REALS for robust and efficient batch image alignment through simultaneous transformation and low rank and sparse decomposition. REALS is constructed upon our finding that the low rank subspace can be recovered via linear projection, which allows us to perform simultaneous image alignment and decomposition with gradient-based updates. REALS achieves orders-of-magnitude improvement in terms of accuracy and speed compared to the state-of-the-art robust image alignment algorithms.

1. Introduction

Calcium imaging has enabled simultaneous recording of the activity of a large population of neurons *in vivo* owing to the recent advancement of genetically-encoded calcium indicators [9, 10, 33, 43] and fluorescence microscopy techniques [1, 6, 11, 29, 39]. State-of-the-art calcium imaging methods allow data acquisition at up to thousands of frames per second [34] which corresponds to gigabytes of data per second [15]. Such a large amount of data has not only provided exciting opportunities for system neuroscience, but also introduced new challenges for the data analysis [15]. General pipelines for processing large scale calcium imaging data typically start with motion correction or image alignment [13], which is critical for accurate extraction of spiking activities from a calcium imaging dataset that suffer from the motion of the brain.

Unfortunately, the neuronal activity itself — the signal of

our interests — in the images makes even the well-aligned images significantly different from one another [20] which poses a challenge to the accurate image alignment. This indicates that alignment needs to be done considering the underlying structure of the images rather than relying solely on the pixel level information.

To address this problem, there have been multiple attempts to employ robust image alignment algorithms such as Robust Alignment by Sparse and Low rank decomposition (RASL) [26], that aims to align a set of significantly different images, for the alignment of calcium imaging data [2, 14, 28]. RASL unified image alignment with low rank and sparse decomposition [7] as a single optimization problem. Through low rank and sparse decomposition, various forms of image corruption, such as noise and occlusion, are absorbed by the sparse component so that the alignment performed on the low rank component is not affected by them. This opened up the possibility of accurate alignment of calcium imaging data.

Unfortunately, the adoption of the robust alignment algorithms has been limited to processing relatively small sized datasets largely due to their high computational costs. The cost mainly comes from singular value decomposition (SVD), which has a computational complexity of $O(nm^2 + n^2m)$ where n and m are the sizes of the data along two dimensions. That is, m is the number of pixels in each image, and n is the number of images in a typical configuration. This indicates that the computational cost increases rapidly with an increase in the data size.

To overcome this limitation, we propose Robust and Efficient ALignment through Simultaneous low rank and sparse decomposition (REALS) for fast and robust image alignment. The main contributions of our work are as follows:

- We provide a mathematical proof, upon which REALS is built, that the low rank subspace can be recovered via linear projection.

---

*Equal contribution

- We propose a robust image alignment algorithm RE-ALS with a computational complexity of $O(nm)$ as opposed to $O(nm^2 + n^2m)$ of the existing methods.

- We propose two extended versions of REALS, multi-resolution REALS and deformable REALS, that achieve even higher alignment accuracy under challenging conditions.

We demonstrate its capability by applying it to the *in vivo* calcium imaging datasets which require robustness, scalability, and computational efficiency.

2. Related Works

2.1. Low rank and sparse decomposition

Candès *et al*. [7] proposed robust principal component analysis (RPCA), which recovers a low rank matrix $L \in \mathbb{R}^{m \times n}$ and a sparse matrix $S \in \mathbb{R}^{m \times n}$ such that $Y = L + S$ for a given data matrix $Y \in \mathbb{R}^{m \times n}$, assuming that the data matrix Y is approximately drawn from a low rank subspace. This is based on the following optimization problem:

$$\min_{L,S} \text{rank}(L) + \lambda||S||_0 \quad \text{subject to} \quad Y = L + S \quad (1)$$

where $||S||_0$ is the L_0 norm of S and $\lambda > 0$ is a hyperparameter. While (1) is known to be computationally intractable, Emmanuel *et al*. [7], and Wright *et al*. [35] showed that the exact solution of (1) can be obtained through the following convex optimization under weak assumptions:

$$\min_{L,S} ||L||_* + \lambda||S||_1 \quad \text{subject to} \quad Y = L + S \quad (2)$$

where $||L||_*$ is the nuclear norm (i.e., the sum of the singular values) of L and $||S||_1$ is the L_1 norm of S. This formulation sparked the development of a series of algorithms for solving (2) [7, 8, 35] and various optimization methods were adopted to reduce the computation time [23, 24, 44, 45]. In addition, online versions of RPCA based on stochastic optimization were introduced to reduce the memory cost [12, 16, 17, 37]. Recently, Han *et al*. [15] proposed an algorithm for efficient approximation of RPCA.

2.2. Robust image alignment

Peng *et al*. [26] formulated a batch image alignment problem as follows: Suppose we have n misaligned and corrupted images $I_1, I_2, \ldots, I_n \in \mathbb{R}^{w \times h}$ of the same object. Assume the corruptions typically affect only a small fraction of all pixels in an image so that they can be modeled as a sparse matrix whose nonzero entries can have arbitrary values. Let s_i represent the error that appears in image I_i. Then, there exist transformations (e.g., affine transformation) $\tau_1, \ldots, \tau_n$ such that the images $\{\tau_i(I_i) - s_i\}_{i=1}^n$ are well aligned to each other and have no corruptions. Since

$\{\tau_i(I_i) - s_i\}_{i=1}^n$ is a set of the aligned images without error, the following matrix L should be approximately low rank:

$$L = [\mathbf{v}(I_1^0)|\cdots|\mathbf{v}(I_n^0)] \in \mathbb{R}^{m \times n} \quad (3)$$

where $I_i^0 = \tau_i(I_i) - s_i$, and $\mathbf{v} : \mathbb{R}^{w \times h} \to \mathbb{R}^m$ ($m = w \times h$) denote the vectorization of an image. Based on this reasoning, the image alignment problem is formulated as the following optimization problem:

$$\min_{L,S,\tau} \text{rank}(L) \quad \text{subject to} \quad \tau(Y) = L + S, \ ||S||_0 \leq k$$
$$(4)$$

where $Y = [\mathbf{v}(I_1)|\cdots|\mathbf{v}(I_n)]$, $S = [\mathbf{v}(s_1)|\cdots|\mathbf{v}(s_n)]$, and $\tau(Y) = [\mathbf{v}(I_1 \circ \tau_1)|\cdots|\mathbf{v}(I_n \circ \tau_n)]$. The Lagrangian form of this problem is as follows:

$$\min_{L,S,\tau} \text{rank}(L) + \gamma||S||_0 \quad \text{subject to} \quad \tau(Y) = L + S \quad (5)$$

where $\gamma > 0$ is a hyperparameter. To handle the nonlinearity of the constraint $\tau(Y) = L + S$, Peng *et al*. [26] approximated the constraint by linearizing the current estimate of τ so that the linearized convex optimization problem can be efficiently solved using the Augmented Lagrange Multiplier method [23]. They relaxed the objective (5) to its convex surrogate and linearized the nonlinear constraint, $\tau(Y) = L + S$, by assuming that the change in τ is small.

Zhang *et al*. [41] generalized the problem in (5) into a low rank tensor recovery framework to exploit spatial structures in images. They minimized the tensor rank, defined as the ranks of a set of unfolding matrices, acquired by applying Tucker decomposition [21] to the given tensor. The cost of solving the optimization problem in both [26, 41] comes mainly from SVD which has a computational complexity of $O(nm^2 + n^2m)$.

t-GRASTA (transformed GRASTA) [18] is based on GRASTA [17] which is a low rank subspace estimation algorithm that represents a low rank matrix as a product of two matrices through gradient descent performed on the Grassmannian. The computational complexity of t-GRASTA is $O(n^2m)$ [16].

3. Proposed Method

3.1. Robust and efficient alignment through simultaneous low rank and sparse decomposition

Building upon [26], we formulate the robust image alignment problem as follows:

$$\min_{L,S,\tau} ||S||_1 \quad \text{subject to} \quad \tau(Y) = L + S, \ \text{rank}(L) \leq l \quad (6)$$

which is obtained by replacing the rank minimization in (5) by the maximum rank constraint on L where the maximum rank l is a small natural number. Then, we use proposition

1 to find a surrogate that is differentiable with respect to the trainable parameters. Below, $\mathcal{C}(A)$ and $<,>$ denote the column space of a matrix A, and the dot product operation, respectively.

Proposition 1. *For given $Y, L, S \in \mathbb{R}^{m \times n}$ where $Y = L + S$, if $\mathcal{C}(L) \cap \mathcal{C}(S) = \{\vec{0}\}$ then there exists $W \in \mathbb{R}^{m \times l}$ such that $L_i = \sum_j < W_j, Y_i > W_j$ where L_i, Y_i, and W_i $(i = 1, ..., n)$ are the i-th columns of L, Y, and W, respectively.*

Proof. Without loss of generality, let $m > n$. Let an arbitrary orthonormal basis of $\mathcal{C}(L)$ be $L_o = \{v_1, ..., v_l\}$. Since $\mathcal{C}(L) \cap \mathcal{C}(S) = \{\vec{0}\}$, there exists $S_0 = \{v_{l+1}, ..., v_{l+s}\}$ that spans $\mathcal{C}(S)$ such that $\{v_1, ..., v_{l+s}\}$ is orthonormal. Note that $l + s < m$. Then, there exists $\{v_{l+s+1}, ..., v_m\}$ such that $\{v_1, ..., v_m\}$ is an orthonormal basis of $\mathbb{R}^m$. Let P be an $m \times m$ matrix such that the j-th column of P^T is v_j $(j = 1, ..., m)$. Then, P is the change-of-basis matrix whose columns are the new coordinate vectors v_j's on the standard basis of $\mathbb{R}^m$. Then, we can construct a diagonal matrix $D \in \mathbb{R}^{m \times m}$ such that $PL = DPY$ by setting $D_{ii} = 1$ if $1 \leq i \leq l$, and $D_{ii} = 0$ otherwise. Note that P is invertible and orthonormal. Since $P^{-1} = P^T$, $L = P^T DPY = (P^T \tilde{D}^T)(\tilde{D}P)Y$ where $\tilde{D} \in \mathbb{R}^{l \times m}$ such that $\tilde{D}_{ii}$ equals 1 if $1 \leq i \leq l$ and 0 otherwise. Therefore, by setting $W = P^T \tilde{D}^T$, $L_i = \sum_j < W_j, Y_i > W_j$. $\square$

Thus, there exists a matrix that projects the matrix Y to its low rank subspace, if $\mathcal{C}(L) \cap \mathcal{C}(S) = \{\vec{0}\}$. This condition is approximately met, if rank(L) is small and S is a sparse matrix (i.e., if S_i, the i-th column of S, is an m-dimensional random sparse vector and rank$(L) \ll m$, then $\frac{<L_i, S_j>}{\|L_i\|\|S_j\|} \approx 0$ for all $i, j \in \{1, ..., n\}$). Note that the same assumption was used for the convex relaxation in (2). Under such condition, the optimization problem in (6) is equivalent to the following:

$$\min_{W, \theta} ||S||_1 \quad \text{where}$$

$$S_i = \tau_i^\theta(Y_i) - \sum_j < W_j, \tau_i^\theta(Y_i) > W_j \quad (7)$$

where τ_i^θ is the transformation applied to the i-th column of Y, parameterized by θ using the differentiable image transformation [19] to make the objective function differentiable. The intuition behind this formulation is that L is a low rank subspace of $\tau(Y)$ [7] and it can be found through linear projection of $\tau(Y)$ to a lower dimensional space defined as the column space of W. We note that similar formulation was employed in [15] to find a surrogate, but without mathematical interpretations or proofs.

Importantly, this objective function is differentiable with respect to all optimization parameters and hence allows gradient-based updates. Therefore, convergence to the local optimum is guaranteed [5]. In addition, each gradient-based update of θ and W has a computational complexity of only $O(nm)$: each τ_i^θ consists of linear summation operations on Y_i [19] that linearly scales with the number of pixels m, and updating W requires only matrix multiplications with $W \in \mathbb{R}^{m \times l}$ and $W^T \in \mathbb{R}^{l \times m}$ [15]. This is in contrast to the complexity of $O(nm^2 + n^2 m)$ [26, 41] or $O(n^2 m)$ [18] of the existing methods.

Furthermore, this framework allows us to handle arbitrary sized data through stochastic gradient descent [31]. As a side benefit, since both forward propagation and back-propagation are possible in REALS, it can be easily combined with other trainable functions for end-to-end learning. As an example, we show that REALS can be combined with non-negative matrix factorization (NMF) [22] to perform unsupervised neuron segmentation in Section 5.3.

It should be noted that the global minima of both (5) and (7) are zeros which can be achieved by any transformation τ^θ that gives transformed images full of zeros. To avoid convergence to such an undesired global minimum, we initialized τ^θ as identity mapping at the beginning of optimization and clamped the values of θ after each parameter update (see Algorithm S1 in the supplementary materials). Instead of parameter clamping, one can set a reference image by freezing the transformation parameters for the image.

REALS can be easily extended to handle 3-D images by extending vectorization $\mathbf{v}$ as $\mathbf{v} : \mathbb{R}^{w \times h \times d} \to \mathbb{R}^m$ $(m = w \times h \times d)$. This extension maintains the differentiability of the objective function of REALS with respect to W and θ.

3.2. Multi-resolution REALS

We introduce an extended version of REALS that makes use of a multi-resolution image pyramid, which is a common strategy for aligning a set of large images with large misalignment [25, 32]. For a set of images $Y = [\mathbf{v}(I_1)|\cdots|\mathbf{v}(I_n)]$ where $I_i \in \mathbb{R}^{w \times h}$, we generate an image pyramid $Y_{(k)} = [\mathbf{v}(I_{1,(k)})|\cdots|\mathbf{v}(I_{n,(k)})]$ where $k \in \mathbb{N}$ and $I_{i,(k)} \in \mathbb{R}^{\frac{w}{2^k} \times \frac{h}{2^k}}$ is a downsampled image of I_i. Then, for each $Y_{(k)}$ $(k = 0, \ldots, K)$, we have an optimization objective as in (7). Since the transformation should remain the same for a given image regardless of k assuming that the translation parameters are defined at a relative scale, the objective function can be written as follows:

$$\min_{\{W_{(i)}\}, \theta} \Sigma_{k=0}^K \beta_k ||S_{(k)}||_1 \quad (8)$$

where $\beta_k \in \mathbb{R}_+$ is a hyperparameter that controls the relative contribution of each level, and $S_{(k)}$ and corresponding $W_{(k)}$ from the image pyramid $Y_{(k)}$ are constructed as in (7). Since this objective function is differentiable with respect to θ and $W_{(k)}$ for all $k \in \{0, \ldots, K\}$, gradient-based optimization can be used to update the parameters.

3.3. Deformable REALS

As the optimization problem in REALS is formulated in a form that allows backpropagation, it can be easily extended to handle affine transformation and deformable transformation simultaneously by setting the transformation τ^θ in (7) as a composite function of parameterized affine and deformable transformations (i.e., $\tau^\theta = \tau^{\theta_1} \circ \tau^{\theta_2}$, where τ^{θ_1} is deformable transformation and τ^{θ_2} is affine transformation). The deformable transformation τ^{θ_1} is parameterized by the displacement field $\theta_1 \in \mathbb{R}^{w \times h \times 2}$ that encodes that relative position of the each pixel before and after the transformation [3, 4]. The optimization objective of deformable REALS is as follows:

$$\min_{W,\theta_1,\theta_2} ||S||_1 \quad + \quad \lambda_1||\theta_1||_1 \quad + \quad \lambda_2\mathcal{L}_{smooth}(\theta_1) \quad (9)$$

where $||\theta_1||_1$ and $\mathcal{L}_{smooth}(\theta_1)$ are the regularization terms employed to introduce preferences for affine transformation over deformable transformation and spatially smooth deformation [3], respectively. λ_1 and λ_2 are hyperparameters. This simplifies the conventional two-step registration pipelines for the biomedical images [3] that sequentially perform affine and deformable transformations to correspond to the movement and the deformation of biological tissues, respectively.

4. Experiments and Results

We compared REALS with RASL [26], t-GRASTA [18], and l_p+ADMM [41]. For testing REALS, RASL, t-GRASTA, and l_p+ADMM on a CPU, we used a PC with an Intel i7-9700K CPU and 128GB of RAM. The source code distributed by the authors was written in MATLAB without GPU acceleration. For testing the performance of REALS on a GPU, we used a PC with an Intel Xeon Silver 4214 CPU, an NVIDIA GeForce RTX 3090 GPU and 128GB of RAM. REALS was implemented using Pytorch.

4.1. Testing on digit, window and Al Gore dataset

To validate REALS, we verified how REALS and other algorithms (RASL, t-GRASTA, and l_p+ADMM) performed on three datasets that were typically used for testing robust image alignment algorithms [18, 26, 30, 31, 36, 41]. Due to the absence of the ground truth aligned images, the alignment results were compared visually. REALS successfully recovered the low rank images (Fig. 1) and all methods yielded qualitatively similar results (Fig. S1)

4.2. Testing on a zebrafish brain dataset

4.2.1 Larval zebrafish brain dataset

For quantification of the performance, we generated and used a synthetic dataset with known geometric perturba-
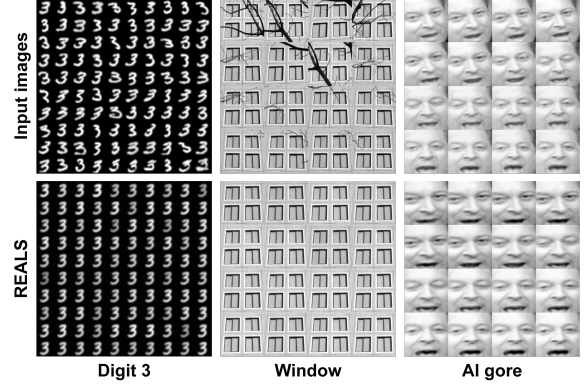

Figure 1: Testing REALS on standard datasets. (Top) Input images. (Bottom) Recovered low rank images.
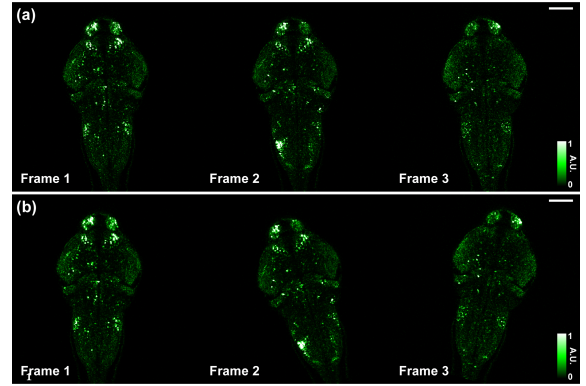

Figure 2: Zebrafish brain dataset. (a) Source images acquired by performing calcium imaging of a larval zebrafish brain. (b) Synthetic images with known geometric perturbations. Scale bar, 100 μm.

tions, from which the alignment accuracy could be measured. We first imaged the whole-brain neuronal activity of a larval zebrafish that was strictly immobilized using both physical (i.e., embedding in a gel) and chemical (i.e., exposing to paralytics) means. Hence, the series of acquired images is free from misalignment while exhibiting significant changes due to neuronal activity and noise as shown in Fig. 2(a). Then, each image is perturbed by Euclidean transformation (i.e., translation and rotation) with random yet known parameters to construct a dataset for testing robust alignment algorithms (Fig. 2(b)). The animal experiments conducted for this study were approved by the Institutional Animal Care and Use Committee (IACUC) of KAIST (KA2021-125).

We note that the acquired images are linear superposition of the underlying structure of the brain and the neuronal activity with time-dependent spatio-temporal patterns, and hence they can be modeled as the sum of a low rank matrix and a sparse matrix [40]. The size of the acquired dataset

was $512(x) \times 512(y) \times 60(t)$. Further details on the dataset are described in the Supplementary Note 1.1.

4.2.2 Performance metric

Previous works relied on human annotated data to estimate alignment accuracy [18, 26, 36] or used the reconstruction error in the recovered low rank matrix as an indirect measure of the alignment accuracy [30, 31, 41], neither of which is suited for the precise assessment of the alignment accuracy on a large scale dataset. As multiple sets of transformations that align the images well with respect to other images are equally valid (i.e., the optimal set of transformations is not unique), we used mean squared error (MSE) defined as follows for evaluating the alignment accuracy:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} ||(\tau_i^{\text{align}} \circ \tau_i^{\text{rand}})(I_i)$$
$$- \mu(\{\tau_j^{\text{rand}} \circ \tau_j^{\text{align}}\}_{j=1}^{n})(I_i)||_F \quad (10)$$

where $\{\tau_i^{\text{rand}}\}_{i=1}^{n}$ is a set of random transformations, $\{\tau_i^{\text{align}}\}_{i=1}^{n}$ is a set of transformations that is acquired with an alignment algorithm, $\mu(\{\tau_j\}_{j=1}^{n})$ is the transformation that averages each parameter from τ_j's $(j = 1, \ldots, n)$, and $\{I_i\}_{i=1}^{n}$ is a set of aligned images (see Movie S1 for the visual assessment of the alignment accuracy with multiple MSE values).

4.2.3 Performance comparison setting

We compared the following methods in terms of MSE and computation time: REALS, RASL [26], t-GRASTA [18], and l_p+ADMM [41]. We used Euclidean transformation for all the methods. We ran REALS for 2000 iterations, and used 1×10^{-4} and 1×10^{-2} as the learning rates for W and τ, respectively. Clamping parameters were set to limit the maximum translation in horizontal and vertical directions as 64 pixels and the maximum rotation angle with respect to the image center as $35°$. The hyperparameters were set as proposed in [18, 26, 41].

4.2.4 Translation and rotation

We first compared the performance of the alignment algorithms on datasets with varying levels of misalignment where the neuronal activity and the native noise in the images served as the outliers. We synthetically perturbed the input images with Euclidean transformations whose angles of rotation were drawn from $U(-\theta_0, \theta_0)$, where $U(a, b)$ denotes a uniform distribution between a and b, and both x and $y-$translations were drawn from $U(-t_0, t_0)$. For RE-ALS, we ran 5 independent trials with random seeds for the synthetic perturbations for the performance measurement at
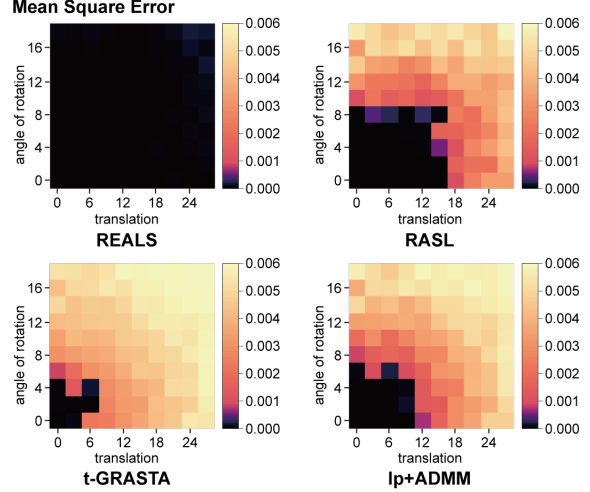

Figure 3: Heatmaps showing the mean squared error of RE-ALS, RASL, t-GRASTA and l_p+ADMM. The algorithms were tested for the datasets with various levels of geometric perturbations (translation and rotation).

each setting (see Table S1 for the hyperparameters). For testing other methods, we ran single trials at each setting because of the long computational time.

Fig. 3 shows the MSE of each method on the datasets with different levels of translation and rotation. REALS outperformed other methods by a large margin in terms of MSE. The average computation time and the average MSE of each method measured from the trials in Fig. 3 are summarized in Table 1. REALS achieved an order of magnitude improvement in terms of both speed and accuracy compared to the other methods (see Fig. S2 for the computation time at each experimental setting). In addition, to monitor the convergence behavior of the algorithms, we measured the MSE with a time stamp after each parameter update. As shown in Fig. 4, REALS on GPU took 100-1000 times shorter than RASL, t-GRASTA, and l_p+ADMM to achieve the same level of MSE. REALS on CPU was about 10 times slower than that on GPU, but it was still significantly faster than other algorithms. The MSE as a function of the number of parameter updates is in Fig. S3 which shows that RE-ALS requires smaller number of iterations for convergence on top of the cheaper complexity of each iteration.

Method	Time ($\pm$std)	Mean squared error
REALS	$\mathbf{79.8 \pm 2.4}$	$\mathbf{3.9 \times 10^{-5}}$
RASL	2419.2 ± 46.2	2.4×10^{-3}
t-GRASTA	856.2 ± 78.6	3.4×10^{-3}
l_p+ADMM	2305.8 ± 4395.6	4.3×10^{-3}

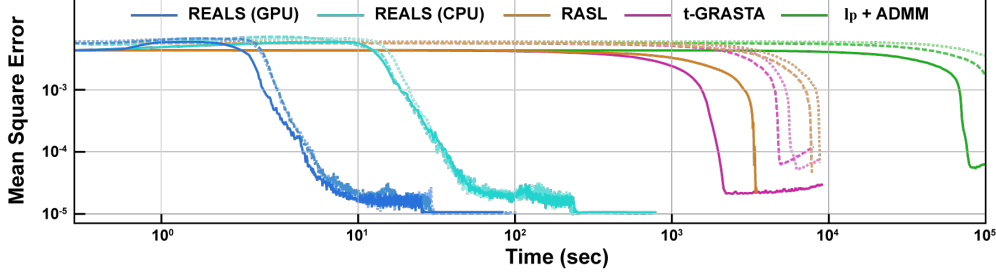Table 1: Computation time (unit: second) and mean squared error of REALS, RASL, t-GRASTA and l_p+ADMM.

Figure 4: Mean squared error of REALS, RASL, t-GRASTA, l_p+ADMM versus computation time. Solid: $(t_0, \theta_0) = (12, 8°)$, Dashed: $(t_0, \theta_0) = (18, 12°)$, Dotted: $(t_0, \theta_0) = (24, 16°)$.

4.2.5 Additional image corruption

To compare the robustness of the alignment algorithms against image corruption, we added Poisson and Gaussian noise to the input images and measured the MSE (see Movie S2 for visual assessment of the noise level). We varied the level of Poisson noise and Gaussian noise and perturbed the images with multiple combinations of translation and rotations: $(t_0, \theta_0) \in \{(0, 0°), (6, 4°), (12, 8°), (18, 12°), (24, 16°)\}$. To add Poisson noise to the images, we first normalized the input images and multiplied them by the brightness level ($\in \{33, 100, 333, 1000, \infty\}$), and then used each pixel value as the parameter (i.e., mean value) of the Poisson distribution. In this configuration, the physical meaning of the brightness level is the expected number of photons that hit the brightest pixel. For Gaussian noise, we first normalized the input images and added Gaussian noise with a standard deviation of $\sigma \in \{0.0, 0.001, 0.003, 0.01, 0.03\}$.

In addition, as another way of testing the robustness of the alignment algorithms, we synthesized the images of neuronal activity with the reduced low rank components — which is equivalent to the relative increase of sparse components — as follows. We first applied BEAR [15] to the larval zebrafish brain dataset Y and acquired low rank matrix L and sparse matrix S. Then, we reconstructed the images with the reduced low rank component as $Y_\alpha = \alpha L + S$, where $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, and then applied random perturbation to Y_α. This setting with $\alpha < 1$ can be considered as simulating future neuronal activity imaging data, as the relative brightness change of the optical neuronal activity indicators compared to their baseline brightness — which determines that ratio of S and L — has been continuously increasing [42].

We measured the MSE for the datasets with these additional image corruptions and the results are reported in Table 2 (see Fig. S4 for the MSE at each experimental setting). REALS outperformed other methods under challenging conditions which proves its robustness against various forms of image corruption.

Method	Poisson noise	Gaussian noise	Reduced low rank
REALS	$\mathbf{3.9 \times 10^{-4}}$	$\mathbf{2.1 \times 10^{-4}}$	$\mathbf{7.7 \times 10^{-5}}$
RASL	2.8×10^{-3}	2.5×10^{-3}	1.8×10^{-3}
t-GRASTA	3.1×10^{-3}	3.0×10^{-3}	2.7×10^{-3}
l_p+ADMM	3.6×10^{-3}	3.5×10^{-3}	3.4×10^{-3}

Table 2: Mean squared error of REALS, RASL, t-GRASTA and l_p+ADMM for the images with additional corruption.

4.2.6 Multi-resolution REALS

(t_0, θ_0)	Input images*	REALS	m-REALS†
$(30, 25°)$	0.0068	0.0013	**0.0002**
$(45, 30°)$	0.0078	0.0019	**0.0004**
$(60, 35°)$	0.0082	0.0024	**0.0005**
$(75, 40°)$	0.0088	0.0056	**0.0019**
$(90, 45°)$	0.0094	0.0077	**0.0049**

Table 3: Mean squared error (MSE) of REALS and multi-resolution REALS. *MSE of the input images (i.e., τ^{align} is an identity transformation). †multi-resolution REALS.

We tested the performance of the multi-resolution REALS under challenging conditions (see Table S2 for the hyperparameters). We first added Poisson noise that corresponded to the brightness level of 33 and Gaussian noise with $\sigma = 0.03$ to the images. We then applied large geometric perturbations that ranged from $(t_0, \theta_0) = (30, 25°)$ to $(t_0, \theta_0) = (90, 45°)$. We ran 5 independent trials to compare REALS and multi-resolution REALS. Multi-resolution REALS showed up to 5 times lower MSE than REALS as summarized in Table 3.

4.2.7 Deformable REALS

We compared the performance of deformable REALS, introduced in Section 3.3, with VoxelMorph [3] using images that were perturbed with deformable transformations. For the experiment, we generated synthetic datasets by se-

(a) Average projection

Misaligned (Input)   Aligned (Ground-truth)   REALS   VoxelMorph

(b) Representative frames (enlarged view)

Input (1st frame)   Input (34th frame)   REALS (34th frame)   VoxelMorph (34th frame)
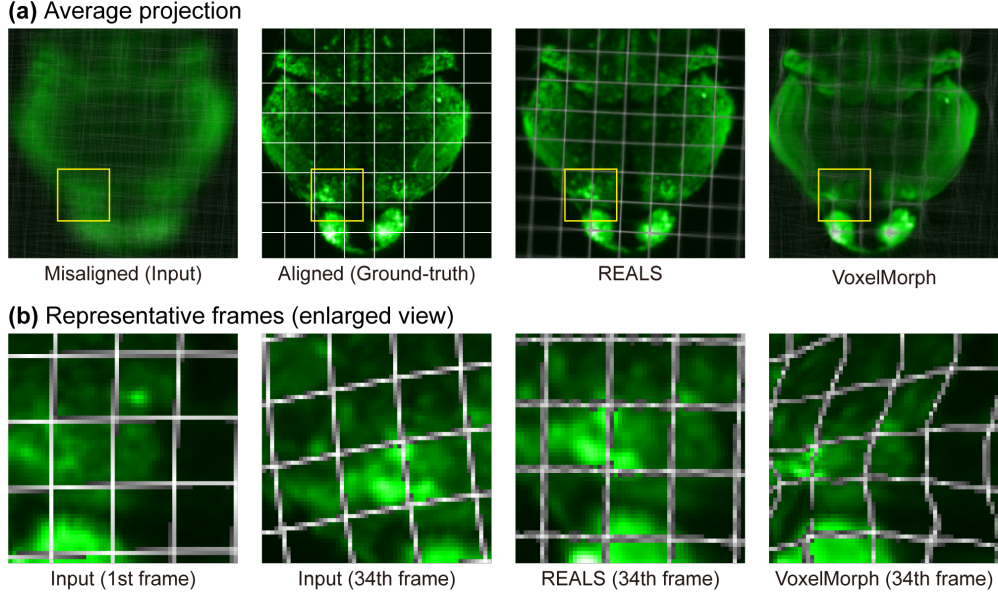
Figure 5: Deformable image registration using deformable REALS and VoxelMorph. (a) Temporal average projections of the misaligned input images, ground truth aligned images, images aligned using deformable REALS, and images aligned using VoxelMorph. Regular grids with a pitch of 32 pixels were overlaid on the ground truth aligned images. Thus, the sharpness of the images and the grids in the average projections shows the alignment accuracy. The grids were not used for the alignment, but overlaid only for visual aid. (b) Representative frames of the input images, images aligned using deformable REALS, and images aligned using VoxelMorph. The images show the boxed area in **a**. Regular grids with a pitch of 16 pixels were overlaid on the ground truth aligned images. That is, the grids in the images show how the images were perturbed and aligned.

quentially applying random deformable and Euclidean perturbations to the larval zebrafish brain dataset. Each entity of the displacement field for the deformable perturbation was first drawn from the normal distribution with a mean value of zero and a standard deviation of σ_1. Thereafter, the displacement field was convovled with a 2-D Gaussian kernel with a standard deviation of σ_2 to mimic the deformation of biological tissues. The following parameters were used for the Euclidean and deformable perturbations: $\sigma_1 \in \{0.2, 1\}$, $\sigma_2 = 40$, and $(t_0, \theta_0) \in \{(0, 0°), (3, 2°), (6, 4°), (9, 12°), (12, 8°)\}$ (see Movie S3 for the visual assessment of the level of perturbations).

Fig. 5 shows the temporal average projections and the representative frames of the input (misaligned) images and the registered images acquired using deformable REALS and VoxelMorph. Note that the first frame of the input images was set as the fixed target (reference) image for training and inferencing with VoxelMorph. VoxelMorph failed to preserve the neuronal structure as the result of overly deforming the images to minimize the difference between the input image and the target image; This demonstrates the importance of considering the underlying structure for the alignment. We measured the MSE for each setting and the results are summarized in Table S3 (see Movie S3 for the comparison of deformable REALS and VoxelMorph).

5. Additional Demonstrations

5.1. Testing on mouse brain dataset

We applied REALS to a mouse brain image dataset with a size of $480(x) \times 752(y) \times 1000(t)$ [27] that suffered from large motion. Such motion artifacts are inevitable in *in vivo* mouse brain imaging even with head fixation, because the position of the brain relative to the skull changes with respiration and blood pulsation of the mouse. Despite the challenges posed by the large motion, low image contrast, and low signal-to-noise ratio, REALS successfully aligned the dataset as shown in Fig. 6 (see Table S4 and Movie S4).

5.2. REALS on 3-D data

We applied 3-D REALS on 3-D time-series images of neural activity of a zebrafish brain with a size of $256(x) \times 512(y) \times 48(z) \times 60(t)$. Similar to the 2-D image dataset, we first imaged whole-brain neuronal activity of a larval zebrafish that was strictly immobilized in 3-D. Then, we synthetically perturbed the images with Euclidean transformation whose angles of rotation around the z-axis were sampled from $U(-2, 2)$ (unit: °), and xyz-translations were sampled from $U(-4, 4)$ (unit: pixel). We used affine transformation and optimized REALS with mini-batches with the batch size of 30 (see Table S5 for the hyperparame-
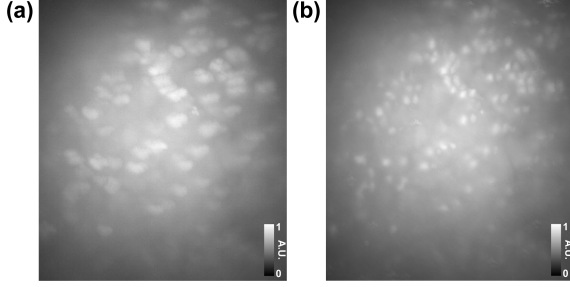
Figure 6: Alignment of neuronal activity images of a mouse brain using REALS. (a) Temporal maximum intensity projection of the input images. (b) Temporal maximum intensity projection of the aligned image.

ters). Fig. 7 shows the temporal maximum intensity projections (MIPs) of the input and aligned images (see Movie S5). Fine neuronal structures are clearly visible in the temporal MIP after the alignment.
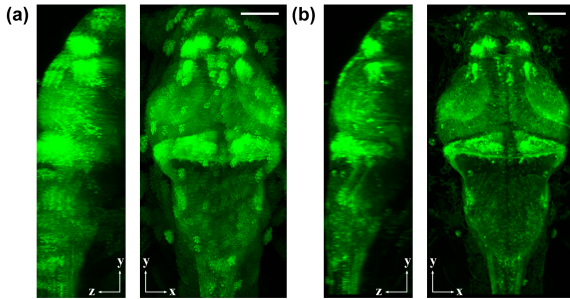

Figure 7: Alignment of 3-D neuronal activity images of a larval zebrafish brain using REALS. (a) Temporal maximum intensity projection of the 3-D input images. (b) Temporal maximum intensity projection of the 3-D aligned images. Scale bar, 100 μm.

5.3. REALS with NMF

We integrated REALS with NMF [22] to simultaneously perform image alignment, separation of neuronal activity from background, and unsupervised segmentation of neurons; In order to retain differentiability with respect to all parameters, we employed projective NMF [38] which is a differentiable formulation of NMF.

We processed a zebrafish calcium imaging video with a size of $256(x) \times 512(y) \times 600(t)$ which suffer from motion artifacts using REALS with NMF, with and without optimizing the transformation parameters. Fig. 8 shows the segmentation result, in which a random color is assigned to each spatial component. The result without optimizing the transformation parameters showed severe over-segmentation due to the motion artifacts (Fig. 8(b)). However, with transformation parameter optimization, confined

spatial components that correspond to the neurons were obtained which manifested as the color uniformity within each neuron in Fig. 8(d) (see Table S6 and Movie S6).
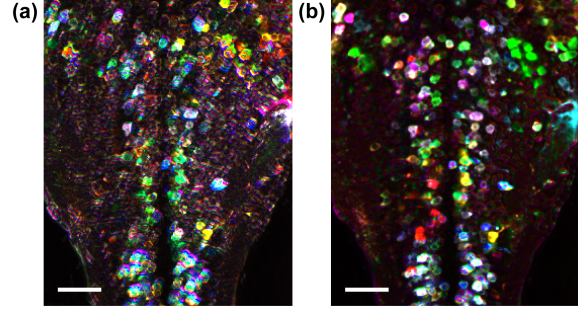

Figure 8: REALS with NMF for simultaneous alignment, low rank and sparse decomposition, and NMF. (a) Spatial footprints from REALS with NMF, but without alignment, are colored and overlaid. (b) Spatial footprints from REALS with NMF are colored and overlaid. Same random color palette as in **a** is used. Scale bar, 30 μm.

6. Conclusion

In this paper, we proposed REALS, an algorithm for robust and efficient batch image alignment. REALS employs a differentiable image transform and differentiable low rank and sparse decomposition which are integrated in a form that we can backpropagate for gradient-based updates. This characteristic comes with multiple benefits. First, it has a significantly reduced computation complexity of $O(nm)$ and achieves orders-of-magnitude of performance improvement. Second, it can be easily generalized for the alignment of high-dimensional image data or deformed images. Third, it can be integrated with other trainable functions that allows backpropagation. The robustness, computational efficiency, and scalability of REALS make it an ideal algorithm for the alignment of calcium imaging data.

Acknowledgement

References

[1] Sara Abrahamsson, Jiji Chen, Bassam Hajj, Sjoerd Stallinga, Alexander Y. Katsov, Jan Wisniewski, Gaku Mizuguchi, Pierre Soule, Florian Mueller, and Claire Dugast Darzacq. Fast multicolor 3d imaging using aberration-corrected multifocus microscopy. *Nature methods*, 10(1):60–63, 2013.

[2] Sophie Aimon, Takeo Katsuki, Tongqiu Jia, Logan Grosenick, Michael Broxton, Karl Deisseroth, Terrence J. Sejnowski, and Ralph J Greenspan. Fast near-whole–brain imaging in adult drosophila during responses to stimuli and behavior. *PLoS biology*, 17(2):e2006732, 2019.

[3] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, Adrian V Dalca, and John Guttag. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9252–9260, 2018.

[4] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.

[5] Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *Siam Review*, 60(2):223–311, 2018.

[6] Matthew B. Bouchard, Venkatakaushik Voleti, César S. Mendes, Clay Lacefield, Wesley B. Grueber, Richard S. Mann, Randy M. Bruno, and Elizabeth Hillman. Swept confocally-aligned planar excitation (scape) microscopy for high-speed volumetric imaging of behaving organisms. *Nature photonics*, 9(2):113–119, 2015.

[7] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *CoRR*, abs/0912.3599, 2009.

[8] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[9] Qian Chen, Joseph Cichon, Wenting Wang, Li Qiu, Seok-Jin R. Lee, Nolan R. Campbell, Nicholas DeStefino, Michael J. Goard, Zhanyan Fu, and Ryohei Yasuda. Imaging neural activity using thy1-gcamp transgenic mice. *Neuron*, 76(2):297–308, 2012.

[10] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, and Vivek Jayaraman. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, 2013.

[11] Lin Cong, Zeguan Wang, Yuming Chai, Wei Hang, Chunfeng Shang, Wenbin Yang, Lu Bai, Jiulin Du, Kai Wang, and Quan Wen. Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (danio rerio). *Elife*, 6, 2017.

[12] Jiashi Feng, Huan Xu, and Shuicheng Yan. Online robust pca via stochastic optimization. In *Advances in Neural Information Processing Systems*, pages 404–412, 2013.

[13] Andrea Giovannucci, Johannes Friedrich, Pat Gunn, Jérémie Kalfon, Brandon L. Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L. Gauthier, and Pengcheng Zhou. Caiman an open source tool for scalable calcium imaging data analysis. *Elife*, 8:e38173, 2019.

[14] Logan Grosenick, Michael Broxton, Christina K. Kim, Conor Liston, Ben Poole, Samuel Yang, Aaron Andalman, Edward Scharff, Noy Cohen, and Ofer Yizhar. Identification of cellular-activity dynamics across large tissue volumes in the mammalian brain. *BioRxiv*, 2017.

[15] Seungjae Han, Eun-Seo Cho, Inkyu Park, Kijung Shin, and Young-Gyu Yoon. Efficient neural network approximation of robust pca for automated analysis of calcium imaging data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 595–604. Springer, 2021.

[16] Jun He, Laura Balzano, and John Lui. Online robust subspace tracking from partial information. *arXiv preprint arXiv:1109.3827*, 2011.

[17] Jun He, Laura Balzano, and Arthur Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1568–1575, 2012.

[18] Jun He, Dejiao Zhang, Laura Balzano, and Tao Tao. Iterative online subspace learning for robust image alignment. In *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2013.

[19] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.

[20] Dal Hyung Kim, Jungsoo Kim, João C Marques, Abhinav Grama, David G C Hildebrand, Wenchao Gu, Jennifer M. Li, and Drew N. Robson. Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nature Methods*, 14(11):1107–1114, 2017.

[21] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[22] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[23] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[24] Zhouchen Lin, Arvind Ganesh, John Wright, Leqin Wu, Minming Chen, and Yi Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *Coordinated Science Laboratory Report no. UILU-ENG-09-2214, DC-246*, 2009.

[25] Abhishek Nan, Matthew Tennant, Uriel Rubin, and Nilanjan Ray. Drmime: Differentiable mutual information and matrix exponential for multi-resolution image registration. In *Medical Imaging with Deep Learning*, pages 527–543, 2020.

[26] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2233–2246, 2012.

[27] Eftychios A. Pnevmatikakis and Andrea Giovannucci. Normcorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of Neuroscience Methods*, 291:83–94, 2017.

[28] Benjamin Poole, Logan Grosenick, Michael Broxton, Karl

Deisseroth, and Surya Ganguli. Robust non-rigid alignment of volumetric calcium imaging data. *COSYNE*, 2015.

[29] Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, and Edward S Boyden. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature methods*, 11(7):727–730, 2014.

[30] Duo Qiu, Minru Bai, Michael K. Ng, and Xiongjun Zhang. Robust Low Transformed Multi-Rank Tensor Methods for Image Alignment. *Journal of Scientific Computing*, 87(1):1–40, 2021.

[31] Wenjie Song, Jianke Zhu, Yang Li, and Chun Chen. Image alignment by online robust pca via stochastic gradient descent. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(7):1241–1250, 2016.

[32] Philippe Thévenaz and Michael Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.

[33] Lin Tian, S. Andrew Hires, Tianyi Mao, Daniel Huber, M. Eugenia Chiappe, Sreekanth H. Chalasani, Leopoldo Petreanu, Jasper Akerboom, Sean A. McKinney, and Eric R. Schreiter. Imaging neural activity in worms, flies and mice with improved gcamp calcium indicators. *Nature methods*, 6(12):875–881, 2009.

[34] Venkatakaushik Voleti, Kripa B. Patel, Wenze Li, Citlali Perez Campos, Srinidhi Bharadwaj, Hang Yu, Caitlin Ford, Malte J. Casper, Richard Wenwei Yan, Wenxuan Liang, Chentao Wen, Koutarou D. Kimura, Kimara L. Targoff, and Elizabeth M C Hillman. Real-time volumetric microscopy of in vivo dynamics and large-scale samples with scape 2.0. *Nature methods*, 16(10):1054–1062, 2019.

[35] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in Neural Information Processing Systems*, pages 2080–2088, 2009.

[36] Yi Wu, Bin Shen, and Haibin Ling. Online robust image alignment via iterative convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1808–1814, 2012.

[37] Wei Xiao, Xiaolin Huang, Fan He, Jorge Silva, Saba Emrani, and Arin Chaudhuri. Online robust principal component analysis with change point detection. *IEEE Transactions on Multimedia*, 22(1):59–68, 2019.

[38] Zhirong Yang, Zhijian Yuan, and Jorma Laaksonen. Projective non-negative matrix factorization with applications to facial image processing. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(08):1353–1362, 2007.

[39] Young-Gyu Yoon, Zeguan Wang, Nikita Pak, Demian Park, Peilun Dai, Jeong Seuk Kang, Ho-Jun Suk, Panagiotis Symvoulidis, Burcu Guner-Ataman, and Kai Wang. Sparse decomposition light-field microscopy for high speed imaging of neuronal activity. *Optica*, 7(10):1457–1468, 2020.

[40] Young-Gyu Yoon, Zeguan Wang, Nikita Pak, Demian Park, Peilun Dai, Jeong Seuk Kang, Ho-Jun Suk, Panagiotis Symvoulidis, Burcu Guner-Ataman, Kai Wang, and Edward S. Boyden. Sparse decomposition light-field mi-

croscopy for high speed imaging of neuronal activity. *Optica*, 7(10):1457–1468, Oct 2020.

[41] Xiaoqin Zhang, Di Wang, Zhengyuan Zhou, and Yi Ma. Robust low-rank tensor recovery with rectification and alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):238–255, 2021.

[42] Yan Zhang, Márton Rózsa, Daniel Bushey, Jihong Zheng, Daniel Reep, Yajie Liang, Gerad Joey Broussard, Arthur Tsang, Getahun Tsegaye, Ronak Patel, Sujatha Narayan, Jing-Xuan Lim, Rongwei Zhang, Misha B. Ahrens, Glenn C. Turner, Samuel S.-H Wang, Karel Svoboda, Wyatt Korff, Erir R. Schreiter, Jeremy P. Hasseman, Ilya Kolb, and Loren L Looger. jgcamp8 fast genetically encoded calcium indicators. *Janelia Research Campus. Online resource.*, 11 2020.

[43] Yan Zhang, Márton Rózsa, Yajie Liang, Daniel Bushey, Ziqiang Wei, Jihong Zheng, Daniel Reep, Gerard Joey Broussard, Arthur Tsang, Getahun Tsegaye, Sujatha Narayan, hristopher J. Obara, Jing-Xuan Lim, Ronak Patel, Rongwei Zhang, Misha B. Ahrens, Glenn C. Turner, Samuel S.-H. Wang, Wyatt L. Korff, Eric R. Schreiter, Karel Svoboda, Jeremy P. Hasseman, Ilya Kolb, and Loren L. Looger. Fast and sensitive gcamp calcium indicators for imaging neural populations. *bioRxiv 2021.11.08.467793*, 2021.

[44] Tianyi Zhou and Dacheng Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the International Conference on Machine Learning*, page 33–40, 2011.

[45] Tianyi Zhou and Dacheng Tao. Greedy bilateral sketch, completion & smoothing. In *Artificial Intelligence and Statistics*, pages 650–658, 2013.