Online Robust Image Alignment via Subspace Learning from Gradient Orientations

Qingqing Zheng[1], Yi Wang[2]*, and Pheng Ann Heng[1]
[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
qqzheng@cse.cuhk.edu.hk, onewang@szu.edu.cn, pheng@cse.cuhk.edu.hk

Abstract

Robust and efficient image alignment remains a challenging task, due to the massiveness of images, great illumination variations between images, partial occlusion and corruption. To address these challenges, we propose an online image alignment method via subspace learning from image gradient orientations (IGO). The proposed method integrates the subspace learning, transformed IGO reconstruction and image alignment into a unified online framework, which is robust for aligning images with severe intensity distortions. Our method is motivated by principal component analysis (PCA) from gradient orientations provides more reliable low-dimensional subspace than that from pixel intensities. Instead of processing in the intensity domain like conventional methods, we seek alignment in the IGO domain such that the aligned IGO of the newly arrived image can be decomposed as the sum of a sparse error and a linear composition of the IGO-PCA basis learned from previously well-aligned ones. The optimization problem is accomplished by an iterative linearization that minimizes the ℓ_1-norm of the sparse error. Furthermore, the IGO-PCA basis is adaptively updated based on incremental thin singular value decomposition (SVD) which takes the shift of IGO mean into consideration. The efficacy of the proposed method is validated on extensive challenging datasets through image alignment and face recognition. Experimental results demonstrate that our algorithm provides more illumination- and occlusion-robust image alignment than state-of-the-art methods do.

1. Introduction

Image alignment is one of the most widely used image processing techniques in computer vision [24]. The technique seeks the optimal image transformations to establish spatial correspondences between different image acquisitions. Applications in video stabilization [19], medical image registration [23], image recognition [27] and visual tracking [31] all leverage alignment to estimate image correspondences. In recent years, with the increasing popularity of the image and video sharing in social networks, such as Facebook and Instagram, we are seeing a dramatic increasing amount of visual data available online. Such enormous data poses great challenges for existing batch image alignment algorithms, due to great illumination variations between images, partial occlusion, gross pixel corruption, and the dynamically increasing images [2]. Therefore, the robust alignment with both memory and time efficiency deems to be a crucial image processing issue to be resolved for handling large and increasing amount of images.

The problem of batch image alignment has been extensively exploited in the literature [5, 15]. Learned-Miller *et al*. [11] minimized a sum of the pixel-stack entropies to align images, and Huang *et al*. [8] further proposed a clustered scale-invariant feature transform (SIFT) based entropy to address the illumination variations involved in different images. Peng *et al*. [18] proposed a robust alignment by sparse and low-rank decomposition (RASL) for a batch of linearly correlated images. In RASL, the optimal transformations are established by exploiting the low-rank property of aligned images. RASL has been widely used for simultaneously aligning multiple images. However, to align a newly arrived image to the previously aligned ones, RASL has to adjust all the previous transformations to seek the matrix rank minimization. In addition, RASL models corruption and occlusion as sparse intensity errors. Nevertheless, many real-world corrupted images contain severe intensity distortions which are dense thus difficult to be subtracted by RASL [13]. In contrast, Li *et al*. [12] arranged the input images into a 3D tensor, and claimed that severe intensity distortions and partial occlusions can be separated out in the gradient and frequency domain. The optimally aligned image tensor is achieved by simultaneously sparsifying a frequency tensor and a gradient error tensor. However, such offline alignment methods are very memory- and

*Corresponding Author

time-consuming, which limits their capability of aligning large and increasing amount of images.

To better address the dynamically increasing images, the online image alignment has become an active research area [17, 21]. Motivated by [18], Wu *et al*. [30] proposed an online robust image alignment (ORIA) method. ORIA decomposes the low-rank component in RASL into the product of a basis matrix and a weight coefficient matrix, where the basis matrix consists of previously well-aligned images. ORIA employs a fixed rank model, and assumes that the aligned image without corruption is a linear composition of well-aligned basis. Although quite efficient on large datasets, the heuristic basis updating scheme using thresholding and replacement reduces the robustness of image alignment. He *et al*. [7] proposed t-GRASTA (transformed Grassmannian Robust Adaptive Subspace Tracking Algorithm) for online image alignment, which updates the basis matrix with a gradient geodesic step on the Grassmannian. Though [7] learns the subspace in Grassmannian, the number of the subspace dimension is set manually and fixed over the whole procedure. Song *et al*. [22] integrated the geometric transformation into the online robust principal component analysis (RPCA) approach for image alignment. In [22], the object function is directly linearized by performing transformations on the recovered noiseless images. The basis matrix is updated using stochastic gradient descent according to each recovered image. However, as well as [18], [7, 22, 30] all assume large errors such as occlusion and corruption among the images are sparse and separable with respect to intensity, which may fail in aligning images with severe intensity distortions.

To address these challenges mentioned above, we propose an online image alignment method via subspace learning from image gradient orientations (IGO). The proposed method is motivated by PCA from gradient orientations provides more reliable low-dimensional subspace than that from pixel intensities (see Figs. 1 and 2 for more details). The principal subspace of corrupted pixel intensities suffers from artifacts (Figs. 2 (a)-(e)), resulting in the poor quality of reconstruction (Figs. 1 (e)-(f)). In contrast, the principal subspace of corrupted gradient orientations appears to be artifact-free (Figs. 2 (f)-(j)), which offers the reconstruction mainly corresponding to the "face" component (Figs. 1 (g)-(h)). Therefore, instead of processing in the intensity domain like conventional methods, we seek alignment in the IGO domain such that the aligned IGO of the newly arrived image can be decomposed as the sum of a sparse error and a linear composition of IGO-PCA basis learned from previously well-aligned ones. We solve the alignment problem efficiently by an iterative linearization that minimizes the ℓ_1-norm of the sparse error. Furthermore, the IGO-PCA basis is adaptively updated based on thin singular value decomposition (SVD) which takes the

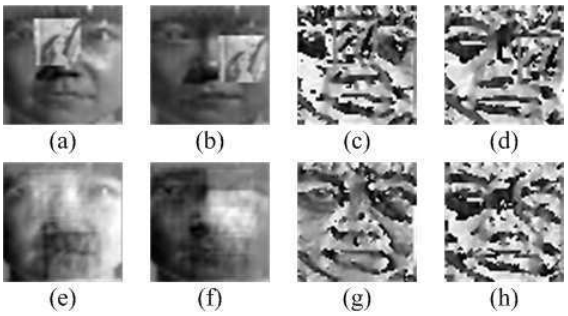

Figure 1: PCA-based reconstruction of pixel intensities and gradient orientations, respectively. (a)-(b) Artificially occluded images from the Yale B database [6]. (c)-(d) Corresponding occluded gradient orientations. (e)-(f) Reconstruction of (a)-(b) with the top five principal components of pixel intensities. (g)-(h) Reconstruction of (c)-(d) with the top five principal components of gradient orientations.

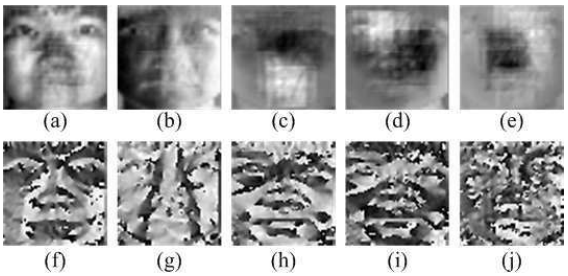

Figure 2: The top five principal components of (a)-(e) pixel intensities and (f)-(j) gradient orientations.

shift of IGO mean into consideration. The benefit of our method is twofold: First, the subspace representation makes our online alignment much more memory-efficient, because we only need to maintain the low-dimensional subspace throughout the whole aligning procedure. Second, the image decomposition and alignment in the IGO domain benefits our method handling the large illumination/occlusion variations that often occur in real-world images, as we show in § 3. We validate the efficacy of the proposed method on extensive challenging datasets through image alignment and face recognition. Experimental results demonstrate that our algorithm provides more illumination- and occlusion-robust image alignment than state-of-the-art methods do.

2. Online robust image alignment

In this section, we first present our robust image alignment algorithm with a known basis, then introduce an efficient method for incrementally updating the basis as new observations arrive. Furthermore, we discuss the memory usage of our algorithm and provide implementation details.

2.1. Robust image alignment in IGO domain

Problem formulation. Suppose we are given n previously aligned grayscale images $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_n \in \mathbb{R}^{w\times h}$ of certain subject, where w and h are width and height of the image. When a new image $\mathbf{I}$ arrives, our task is to seek an optimal transformation $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that warps this image with the previously aligned images.

Conventional methods treat this task as seeking alignment in the image intensity domain such that the newly aligned image can be decomposed as the sum of a sparse error, and a linear composition of either a subset of previously aligned images [30] or a low-rank subspace [7]. The alignment performances of [7, 30] heavily depend on the quality of basis images or the subspace estimated from them. Moreover, both [7, 30] assume large errors such as occlusion and corruption among the images are sparse. However, in many real-world applications, the quality of input images are corrupted by spatially varying intensity distortions, which leads the subspace estimated from the pixel intensities arbitrarily biased (see Figs. 2 (a)-(e)). In contrast, PCA from image gradient orientations (IGO) [25] is able to provide more reliable low-dimensional subspace than that from pixel intensities (see Figs. 2). The IGO-PCA is statistically verified and further applied directly for face reconstruction in [25].

Inspired by [25], instead of processing in the intensity domain, we seek alignment in the IGO domain such that the aligned IGO of the newly arrived image can be decomposed as the sum of a sparse error and a linear composition of IGO-PCA basis learned from previously well-aligned ones. We propose a robust incremental alignment method and formulate it into a constrained ℓ_1 minimization problem as

$$\min_{\mathbf{w},\mathbf{e},\tau} ||\mathbf{e}||_1 \qquad s.t.\ \ vec(\mathbf{\Phi}\circ\tau) = \mathbf{U}\mathbf{w} + \mathbf{e}, \tag{1}$$

where $\mathbf{\Phi} \in [0, 2\pi)^{w\times h}$ is the IGO of the input image $\mathbf{I}$, $\mathbf{U} \in \mathbb{R}^{d\times k}(d = w\times h)$ is a low-rank orthonormal basis estimated from previously well-aligned IGO, $\mathbf{w} \in \mathbb{R}^k$ is the reconstruction weight, and $\mathbf{e} \in \mathbb{R}^d$ measures the dissimilarity between the warped $\mathbf{\Phi}$ and the reconstructed gradient orientation with the subspace $\mathbf{U}$. We denote $vec : \mathbb{R}^{w\times h} \rightarrow \mathbb{R}^d$ as the vectorization operator that stacks a matrix into a vector.

To compute $\mathbf{\Phi}$, we first estimate the image gradients with $\mathbf{G}_w = h_x * \mathbf{I}$ and $\mathbf{G}_h = h_y * \mathbf{I}$. Here h_x and h_y are filters used to approximate the differentiation operator along the image horizontal and vertical direction, respectively [1]. Then we compute the gradient orientation with

$$\mathbf{\Phi} = \arctan(\mathbf{G}_w/\mathbf{G}_h). \tag{2}$$

[1]Possible h_x and h_y can choose Sobel, Prewitt gradient operator, central difference estimator or discrete approximation to the first derivative of the Gaussian.

ADMM solver for linearized convex optimization. The optimization problem in (1) is non-convex and difficult to solve directly due to the nonlinearity of the transformation τ. To tackle this problem, we linearize the constraint by using the local first order Taylor approximation for each IGO as $\mathbf{\Phi}\circ(\tau+\Delta\tau) \approx \mathbf{\Phi}\circ\tau + \mathbf{J}\Delta\tau$, where $\Delta\tau \in \mathbb{R}^p$ is defined by p parameters and $\mathbf{J} = \frac{\partial}{\partial\zeta}(vec(\mathbf{\Phi}\circ\zeta))|_{\zeta=\tau} \in \mathbb{R}^{d\times p}$ is the Jacobian of $\mathbf{\Phi}$ with respect to the transformation τ. Thus (1) can be relaxed into a convex optimization as

$$\begin{aligned} \min_{\mathbf{w},\mathbf{e},\Delta\tau} \quad & ||\mathbf{e}||_1 \\ s.t. \quad & vec(\mathbf{\Phi}\circ\tau) + \mathbf{J}\Delta\tau = \mathbf{U}\mathbf{w} + \mathbf{e}. \end{aligned} \tag{3}$$

In this way, the resulting convex programming in (3) can be efficiently solved by augmented Lagrangian multiplier (ALM) method [4]. Specifically, we formulate the following augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}(\mathbf{w},\mathbf{e},\Delta\tau,\mathbf{y},\mu) =& ||\mathbf{e}||_1 + \mathbf{y}^T f(\mathbf{w},\mathbf{e},\Delta\tau) \\ &+ \frac{\mu}{2}||f(\mathbf{w},\mathbf{e},\Delta\tau)||_2^2, \end{aligned} \tag{4}$$

where $f(\mathbf{w},\mathbf{e},\Delta\tau) = vec(\mathbf{\Phi}\circ\tau) + \mathbf{J}\Delta\tau - \mathbf{U}\mathbf{w} - \mathbf{e}$; $\mathbf{y} \in \mathbb{R}^d$ is the Lagrangian multiplier and μ is a positive hyperparameter.

Given the current estimated transformation τ, the Jacobian matrix $\mathbf{J}$ and the subspace $\mathbf{U}$, each of the subproblem in (4) can be decoupled by the alternating direction method of multipliers (ADMM) [3] and calculated as follows:

$$\begin{aligned} \mathbf{e}^{t+1} &= \mathcal{S}_{\frac{1}{\mu}}(vec(\mathbf{\Phi}\circ\tau) + \mathbf{J}\Delta\tau^t - \mathbf{U}\mathbf{w}^t + \frac{\mathbf{y}^t}{\mu^t}), \\ \mathbf{w}^{t+1} &= \mathbf{U}^T(vec(\mathbf{\Phi}\circ\tau) + \mathbf{J}\Delta\tau^t - \mathbf{e}^{t+1} + \frac{\mathbf{y}^t}{\mu^t}), \\ \Delta\tau^{t+1} &= \mathbf{J}^\dagger(\mathbf{U}\mathbf{w}^{t+1} + \mathbf{e}^{t+1} - vec(\mathbf{\Phi}\circ\tau) - \frac{\mathbf{y}^t}{\mu^t}), \\ \mathbf{y}^{t+1} &= \mathbf{y}^t + \mu^t f(\mathbf{w}^{t+1}, \mathbf{e}^{t+1}, \Delta\tau^{t+1}), \\ \mu^{t+1} &= \rho\mu^t, \end{aligned} \tag{5}$$

where $\mathbf{J}^\dagger$ denotes the Moore-Penrose pseudoinverse of $\mathbf{J}$; $\mathcal{S}_{\frac{1}{\mu}}(\mathbf{x}) = \{[\mathbf{x} - \frac{1}{\mu}]_+ - [-\mathbf{x} - \frac{1}{\mu}]_+\}$ is the soft thresholding operator [16], where $[\cdot]_+ = max(\cdot, 0)$; $\rho > 1$ is a penalty constant which monotonically increases $\{\mu^t\}$ to speed up the convergence of the whole algorithm; and superscript t denotes the iteration. We summarize the ADMM solver for (3) in Algorithm 1.

2.2. Online subspace update

Before describing our online subspace update strategy, we briefly demonstrate why $\mathbf{U}$ is the principal component of the low-rank part of aligned IGO. Now suppose we are given n well-aligned IGO images and their corresponding $\{\mathbf{w}_i, \mathbf{e}_i, \Delta\tau_i\}_{i=1}^n$, the optimal subspace $\mathbf{U}$ that satisfies the

Algorithm 1: ADMM Solver for the local convex problem in (3)

Input : An orthogonal basis $\mathbf{U} \in \mathbb{R}^{d\times k}$, a warped and vectorized gradient orientation $\mathbf{x} = vec(\mathbf{\Phi} \circ \tau) \in \mathbb{R}^d$, the corresponding Jacobian matrix $\mathbf{J} \in \mathbb{R}^{d\times p}$, the penalty constant ρ, the tolerance ϵ and maximal iteration $maxIter$.

Output: The weight vector $\mathbf{w} \in \mathbb{R}^k$, the sparse outliers $\mathbf{e} \in \mathbb{R}^d$, locally linearized parameter $\Delta\tau \in \mathbb{R}^p$ and dual vector $\mathbf{y} \in \mathbb{R}^d$

Initialize: $\mathbf{w}^0 = \mathbf{0}, \Delta\tau^0 = \mathbf{0}, \mathbf{y} = \mathbf{0}, \mu = 1$

while $t \leq maxIter$ **do**

  Update sparse part $\mathbf{e}$:

  $\mathbf{e}^{t+1} = \mathcal{S}_{\frac{1}{\mu}}(\mathbf{x} + \mathbf{J}\Delta\tau^t - \mathbf{U}\mathbf{w}^t + \frac{\mathbf{y}^t}{\mu^t})$

  Update weight $\mathbf{w}$:

  $\mathbf{w}^{t+1} = \mathbf{U}^T(\mathbf{x} + \mathbf{J}\Delta\tau^t - \mathbf{e}^{t+1} + \frac{\mathbf{y}^t}{\mu^t})$

  Update $\Delta\tau$:

  $\Delta\tau^{t+1} = \mathbf{J}^\dagger(\mathbf{U}\mathbf{w}^{t+1} + \mathbf{e}^{t+1} - \mathbf{x} - \frac{\mathbf{y}^t}{\mu^t})$

  Update $\mathbf{y}$: $\mathbf{y}^{t+1} = \mathbf{y}^t + \mu^t f(\mathbf{w}^{t+1}, \mathbf{e}^{t+1}, \Delta\tau^{t+1})$

  Update μ: $\mu^{t+1} = \rho\mu^t$

  if $||f(\mathbf{w}^{t+1}, \mathbf{e}^{t+1}, \Delta\tau^{t+1})||_2 \leq \epsilon$ **then**

    Converge and break

  end

end

return $\mathbf{w} = \mathbf{w}^{t+1}, \mathbf{e} = \mathbf{e}^{t+1}, \Delta\tau = \Delta\tau^{t+1}$ *and* $\mathbf{y} = \mathbf{y}^{t+1}$

constraint of (3) can be formulated as a ℓ_2 norm loss minimization problem

$$\min_{\mathbf{U}} \sum_{i=1}^{n} ||vec(\mathbf{\Phi}_i \circ \tau_i) + \mathbf{J}_i\Delta\tau_i - \mathbf{U}\mathbf{w}_i - \mathbf{e}_i||_2. \quad (6)$$

Let $\mathbf{R}_i = vec(\mathbf{\Phi}_i \circ \tau_i) + \mathbf{J}_i\Delta\tau_i - \mathbf{e}_i$ denote the low-rank part of a well-aligned IGO (here $\mathbf{R}_i$ is short for $\mathbf{R}_{(:,i)}, i \in [1, n]$). Thus the problem of identifying the best $\mathbf{U}$ is equivalent to figure out the principle components of $\mathbf{R} \in \mathbb{R}^{d\times n}$ [9], which can be efficiently solved by SVD. Note that here we remove the sparse errors before performing PCA, and hence leads to more reliable estimation of basis $\mathbf{U}$.

Once $m (m \geq 1)$ new images are aligned with the current basis $\mathbf{U}$, we can obtain the corresponding $\mathbf{R}_{(:,n+j)}, (j \in [1, m])$. Then we incrementally update $\mathbf{U}$ by using $\mathbf{R}_{(:,n+j)}$ and some previously stored basis-related variables. Many algorithms have been developed to efficiently update the basis vectors as new data arrive [7, 22]. Most of them assume the sample mean is zero or fixed when updating the eigenbasis. However, in most of real-world applications, the sample mean may change over time as new samples come. To tackle this problem, we take the sample mean of the new data into consideration for subspace update. Different from [20], our subspace update also works as only one sample arrives. To consider the shift of sample mean, we have the following Lemma:

Lemma 1. *Let* $\mathbf{A} = \mathbf{R}_{(:,1:n)}$, $\mathbf{B} = \mathbf{R}_{(:,n+1:n+m)}$ *be the low-rank estimated data matrices and* $\mathbf{C} = [\mathbf{A}\ \mathbf{B}]$ *be a concatenated matrix of all estimated data. Let scatter matrix be the outer product of the centered data matrix. The means and scatter matrices of* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *are* $\bar{\mathbf{R}}_A, \bar{\mathbf{R}}_B, \bar{\mathbf{R}}_C$ *and* $\mathbf{S}_A, \mathbf{S}_B, \mathbf{S}_C$*, respectively. It can be shown that*

$$\begin{aligned} \bar{\mathbf{R}}_C &= \frac{n}{n+m}\bar{\mathbf{R}}_A + \frac{m}{n+m}\bar{\mathbf{R}}_B, \\ \mathbf{S}_C &= \mathbf{S}_A + \sum_{i=n+1}^{n+m}(\mathbf{R}_i - \bar{\mathbf{R}}_C)(\mathbf{R}_i - \bar{\mathbf{R}}_C)^T \\ &\quad + \frac{nm^2}{(n+m)^2}(\bar{\mathbf{R}}_A - \bar{\mathbf{R}}_B)(\bar{\mathbf{R}}_A - \bar{\mathbf{R}}_B)^T. \end{aligned} \quad (7)$$

From Lemma 1, we obtain that the SVD of $(\mathbf{C} - \bar{\mathbf{R}}_C)$ is equal to the SVD of horizontal concatenation of $(\mathbf{A} - \bar{\mathbf{R}}_A\mathbf{1}_{1\times n})$, $(\mathbf{B} - \bar{\mathbf{R}}_C\mathbf{1}_{1\times m})$ and one additional vector $\frac{\sqrt{nm}}{n+m}(\bar{\mathbf{R}}_A - \bar{\mathbf{R}}_B)$, where $\mathbf{1}_{1\times n}$ is a row vector with all elements equal to 1. Suppose $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the task is to compute the SVD of the concatenation of $[\mathbf{A}\ \hat{\mathbf{B}}] = \mathbf{U}^*\mathbf{\Sigma}^*\mathbf{V}^{*T}$, where $\hat{\mathbf{B}} = [(\mathbf{R}_{n+1} - \bar{\mathbf{R}}_C)| \cdots |(\mathbf{R}_{n+m} - \bar{\mathbf{R}}_C)|\frac{\sqrt{nm}}{n+m}(\bar{\mathbf{R}}_A - \bar{\mathbf{R}}_B)]$. Let $\tilde{\mathbf{B}}$ denote the Gram-Schmidt orthonormalization of $\hat{\mathbf{B}}$ with respect to $\mathbf{U}$, namely $\tilde{\mathbf{B}} = orth(\hat{\mathbf{B}} - \mathbf{U}\mathbf{U}^T\hat{\mathbf{B}})$, we can obtain the update of subspace with the following Lemma:

Lemma 2. *Let* $\mathbf{Q} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{U}^T\hat{\mathbf{B}} \\ \mathbf{0} & \tilde{\mathbf{B}}^T\hat{\mathbf{B}} \end{bmatrix}$, $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$ *denotes the thin SVD of* $\mathbf{Q}$*, the update of basis matrix and eigenvalues can be calculated with* $\mathbf{U}^* = [\mathbf{U}\ \tilde{\mathbf{B}}]\tilde{\mathbf{U}}$ *and* $\mathbf{\Sigma}^* = \tilde{\mathbf{\Sigma}}$.

The proof of above lemmas appear in the supplementary.

For completeness, we summarize our method for a newly arrived image in Algorithm 2.

2.3. Analysis of memory usage

We compare the memory usage between the proposed method and RASL with a sequence of d-pixel images. When the N_{th} image comes, RASL needs to store all the unaligned images $\mathbf{D}$, aligned images $\mathbf{D} \circ \tau$, the low-rank component $\mathbf{A}$, the sparse error $\mathbf{E}$ and the Lagrangian multiplier $\mathbf{Y}$, each of which requires the memory of size dN; when calculating $\mathbf{A}$, we assume RASL uses a thin SVD with k singular values, which suppose to store $dk + k^2 + Nk$ elements; for N Jacobian matrix $\mathbf{J}_i$, $\Delta\tau_i$ and τ_i, each needs dp, p and q elements, respectively. Thus RASL requires memory of size $(5 + p)dN + dk + (k + p + q)N + k^2$.

To be fairly compared, we assume the proposed method uses k basis vectors in the SVD procedure. If we process

Algorithm 2: Online Robust Image Alignment from IGO

Input : An initial orthogonal basis $\mathbf{U} \in \mathbb{R}^{d\times k}$ and the corresponding eigenvalue matrix $\mathbf{\Sigma}$, a new unaligned image $\mathbf{I}$ and the corresponding initial transformation τ, filters for difference operator h_x, h_y, and the maximal iteration K

Output: The updated subspace $\mathbf{U}^*$ and the transformation τ^* for the well-aligned image.

Calculate the horizontal and vertical image gradient $\mathbf{G}_w = h_x * \mathbf{I}$ and $\mathbf{G}_h = h_y * \mathbf{I}$, respectively

Calculate the gradient orientation for the input image: $\mathbf{\Phi} = \arctan(\mathbf{G}_w / \mathbf{G}_h)$

while *not converged and iter* $\leq K$ **do**

  Calculate the Jacobian matrix of $\mathbf{\Phi}$:$\mathbf{J} = \frac{\partial(\mathbf{\Phi}\circ\zeta)}{\partial\zeta}|_{\zeta=\tau}$

  Update the warped, normalized and vectorized image with $vec(\mathbf{\Phi}\circ\tau) = \frac{vec(\mathbf{\Phi}\circ\tau)}{||vec(\mathbf{\Phi}\circ\tau)||_2}$

  Estimate the weight $\mathbf{w}$, the locally linearized transformation parameter $\Delta\tau$, sparse outliers $\mathbf{e}$ and dual parameter $\mathbf{y}$ with (5) via Algorithm 1

  Update the eigenbasis $\mathbf{U}^*$, eigenvalues $\mathbf{\Sigma}^*$ and mean $\bar{\mathbf{R}}_C$ according to § 2.2

  Update the transformation parameter: $\tau^* \leftarrow \tau + \Delta\tau$

end

return $\mathbf{U}^*, \tau^*$

one image per time, when the N_{th} image arrives, the proposed method needs to store ($\mathbf{I}$,$\mathbf{G}_w$,$\mathbf{G}_h$,$\mathbf{\Phi}$) for IGO computation, using d elements for each parameter; for image alignment procedure in Algorithm 1, our method stores parameters ($\mathbf{e}$, $\mathbf{y}$, $\mathbf{J}$, $\mathbf{\Phi}\circ\tau$, τ, $\Delta\tau$, $\mathbf{w}$ and $\mathbf{U}$), which requires memory of $(3+p+k)d + p + q + k$ elements. Finally for the subspace update, our method requires to store an additional parameter group ($\bar{\mathbf{R}}_A$, $\bar{\mathbf{R}}_B$, $\bar{\mathbf{R}}_C$, $\tilde{\mathbf{B}}$, $\hat{\mathbf{B}}$, $\mathbf{Q}$, $\mathbf{U}^*$ and $\mathbf{\Sigma}^*$), which also requires $7d + (k+2)^2 + dk + k^2$ elements. Therefore, the total memory required for our method is $(14+p+2k)d + 2k^2 + 5k + p + q + 4$.

It shows that the memory required by RASL scales with N, while ours keeps constant. For 100 misaligned images, each with 100×100 pixels, assuming $k = 5$ and transformation $\mathbb{G} = Aff(2)$, our method only uses 3% of the memory of RASL while processing the last image. The ratio is even smaller for problems of larger size.

2.4. Implementation detail

In Algorithm 1, we set $\rho = 2$, $\epsilon = 10^{-7}$ and $maxIter = 100$, respectively. In Algorithm 2, the affine transformation $\mathbb{G} = Aff(2)$ is used. To compute the IGO, we use Sobel gradient operator. To get a trained orthonormal basis $\mathbf{U}$ for input, we first perform RASL on a small batch of images, say ten, to obtain the aligned images. Then we estimate the PCA of the aligned images in IGO domain via thin SVD to obtain $\mathbf{U}$ and $\mathbf{\Sigma}$ as [25]. The number of principal components k is automatically determined by an energy ratio (95% in our experiments) of the sum of singular values in $\mathbf{\Sigma}$. The whole algorithm terminates if the number of iterations reach a maximum ($K = 100$) or once the difference of cost between two consecutive iterations is smaller than 10^{-7}. In step 6, we update the basis when the input IGO is well aligned to the current basis, i.e., with the small reconstruction errors $||\mathbf{e}||_2 \leq \delta$ for certain threshold $\delta = 1$.

3. Experiments

In this section, we validate our method on extensive challenging datasets for many real-world applications. To demonstrate the efficacy and robustness of our method, we further compare the performance of three state-of-the-art methods: SIFT feature-based alignment [14], RASL [18] and t-GRASTA [7].

3.1. Image alignment

We first demonstrate the utility of our method for incrementally aligning video sequence *gore*, which contains 140 frames of Al Gore's facial images obtained by a face detector. Fig. 3 shows the alignment results of 20 uniformly sampled frames from the *gore* dataset. It can be observed that our method can generate quite stable alignment on sequential images.

We also test our algorithm on more challenging and unconstrained images taken from the Labeled Faces in the Wild (LFW) [10] dataset, which exhibit great variations in pose and changes in illumination and occlusion. We choose 19 subjects from LFW dataset and each has 35 images. The images are initially cropped to an 80×60 facial frame by applying the common Viola-Jones face detector [26]. We verify the alignment performance by visualizing the average face before and after alignment. We further compare the standard derivation (STD) of selected landmarks' coordinates after alignment. A perfect STD would return a value of zero. For each subject, we annotate the tip of the nose, left eye and right eye from original unaligned images as landmarks.

Fig. 4 visualizes the alignment results by our method, as well as the alignments by SIFT, RASL and t-GRASTA for comparisons. It can be observed that the average faces after alignment by our method is significantly sharper than those before alignment. Furthermore, our method achieves much clearer alignment performance than t-GRASTA and SIFT do, especially in the regions of eyes and mouths (see red and green boxes in Fig. 4). Fig. 5 shows the average X/Y-direction STD results for each subject, respectively. It shows the STD of our method is constantly smaller than

Figure 3: Alignment results of video sequence *gore*. First row: 20 uniformly sampled frames from the 140-frame *gore*. Second row: corresponding aligned frames by our method.

Figure 4: Alignment results of images from LFW. The average faces of different subjects (a) before alignment, (b) after alignment by our method, (c) by SIFT , (d) by RASL and (e) by t-GRASTA. Our average images are much clearer than SIFT's and t-GRASTA's, especially those marked by red and green boxes. (We encourage you to zoom in for better visualization.)

those of t-GRASTA and SIFT, which further demonstrates the superior performance of our method over t-GRASTA and SIFT. These results indicate that the subspace learning from image gradient orientations and basis update strategy embedded in our method contribute to the improvement of image alignment. Meanwhile, our method has nearly the same alignment performance as RASL, see Figs. 4 and 5. However, RASL has a high computational cost for sequential data. On a workstation with Intel Xeon E5-1620 3.70 GHz CPU and 16.0 GB RAM, and with a Matlab implementation, our method averagely spends 0.7 seconds to align a newly arrived image, while RASL averagely needs 673.6 seconds to incrementally align all the images for each subject (images are input into RASL one by one), or 19.2 seconds per image. Moreover, our method is much more memory-efficient than RASL, as discussed in § 2.3.

3.2. Occlusion and illumination variation

To illustrate the robustness of our method, we perform extensive experiments on challenging datasets with occlusion and illumination variation. We first verify our method on dataset Dummy in [18]. We select 30 of 100 images from Dummy, which appear illumination variation and all have artificially added occlusions, to conduct alignment. Fig. 6 visualizes 6 original unaligned images, as well as the aligned images by different methods. Our method together with RASL can successfully align the occluded images, while t-GRASTA or SIFT fails to provide robust alignment.

We further add artificial shadows on images from LFW dataset and test our method on these images. All added shadows are generated randomly and are of various shapes and intensities. Fig. 7 shows some alignment results of a specific subject - Gloria Macapagal Arroyo. The unaligned images are shown in Fig. 7 (a), and the images marked with red boxes are artificially occluded with shadows. The alignment results by our method, as well as the alignments by SIFT, RASL and t-GRASTA, are shown in Figs. 7 (b)-(e), respectively. It is obviously that our method achieves overall best alignment, especially for those images with shadows, whereas neither SIFT, RASL nor t-GRASTA can well handle images with shadows. SIFT is unstable across large occlusion and illumination variations, thus not robust to handle images with various intensity changes. RASL and t-GRASTA assume that the large errors caused by shadows or occlusions among the images are sparse, hence they can then conduct alignment by exploiting the low-rank property

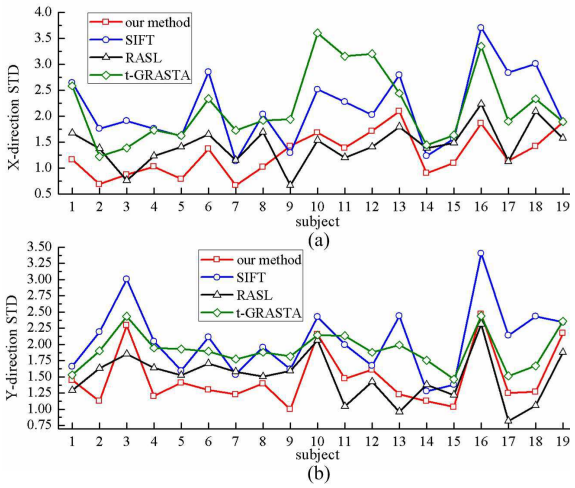

Figure 5: The STD (in pixels) of selected landmarks after alignment. (a) The X-direction and (b) Y-direction STDs.

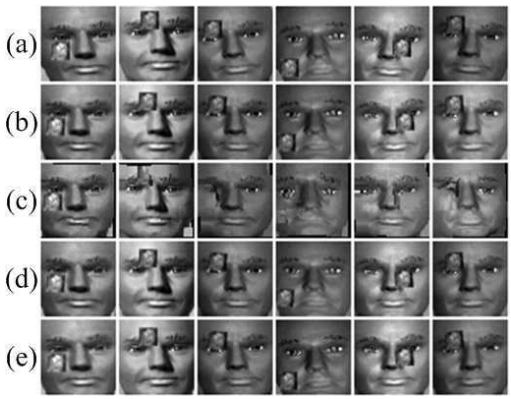

Figure 6: Alignment results of Dummy faces. (a) Original unaligned images. Aligned images (b) by our method, (c) by SIFT, (d) by RASL and (e) by t-GRASTA.

of aligned images. However, many real-world images contain severe intensity distortions, which are not sparse and difficult to be separated. Therefore, RASL and t-GRASTA may fail to handle an image with severe intensity distortion or a batch of images with various shadows or partial occlusions. In contrast, unlike conventional methods [7, 18], we exploit the low-rank property of aligned images in IGO domain, where the aligned IGO of the input image can be truly decomposed as the sum of a sparse error and a linear composition of IGO-PCA basis. In such a way, the intensity distortions presented in the IGO domain are well separated by reconstructing the aligned IGO using reliable IGO-PCA basis. Hence, our method is more practical than [7, 18] in handling real-world applications. It is worth noting that the last input image in Fig. 7 (a) contains natural shadow in face, all comparative methods fail in aligning it well whereas our method aligns it with other images robustly.

3.3. Application to anatomical atlas construction

Image alignment is a fundamental task in medical image processing. Clinically, one of its most important applications is to construct population-based atlas to study anatomical variability [23]. However, robust medical image alignment for atlas construction remains a challenging task, due to the large and incremental amount of images and their diverse intensity distributions caused by different imaging setting. To demonstrate the feasibility of handling complicated medical images, we test our method on a dataset containing 28 magnetic resonance (MR) prostate images [28]. Fig. 8 visualizes the alignment results by different methods. It can be observed that the average image of our method has much clearer prostate boundary than other methods do, especially in the blue boxes shown in Fig. 8 (c). This result demonstrates that our method can provide satisfactory alignment on complicated medical images despite their intensity variations, which is helpful for the prostate atlas construction.

3.4. Application to face recognition

One of the most significant issues in face recognition is how to address pose variations. To tackle this problem, one common way is to align images to a canonical pose prior to recognition step. In this experiment, we employ the sparse representation-based classification (SRC) method [29] for face recognition. Although SRC has achieved impressive recognition performance on public datasets, it does not deal with misalignment between training and testing images, thus is sensitive to the pixel-level misalignment between images. Here we use SRC to evaluate the efficacy of our alignment method and further demonstrate our method is beneficial to practical face recognition applications.

We again use the challenging LFW dataset for this face recognition experiment. For each subject, we randomly select 20 images for training and the other 15 for testing. Because the training images themselves are misaligned, we first employ our method to align them to a canonical frame. With the aligned training images for each subject, a test image is then aligned to training set for recognition by SRC. For comparison purpose, SIFT, RASL, and t-GRASTA are also first applied to align training images, then align test images to training set for recognition. Note that because RASL runs in a batch fashion, we only use it to align training images, and use [27] to align test images to training set.

The face recognition rates of SRC using the four alignment methods are listed in Table 1. The best SRC performance is achieved using our robust alignment method, which demonstrates that SRC benefits from using our alignment as the input. We further compare our result with those listed in a newly published survey on LFW dataset [10].

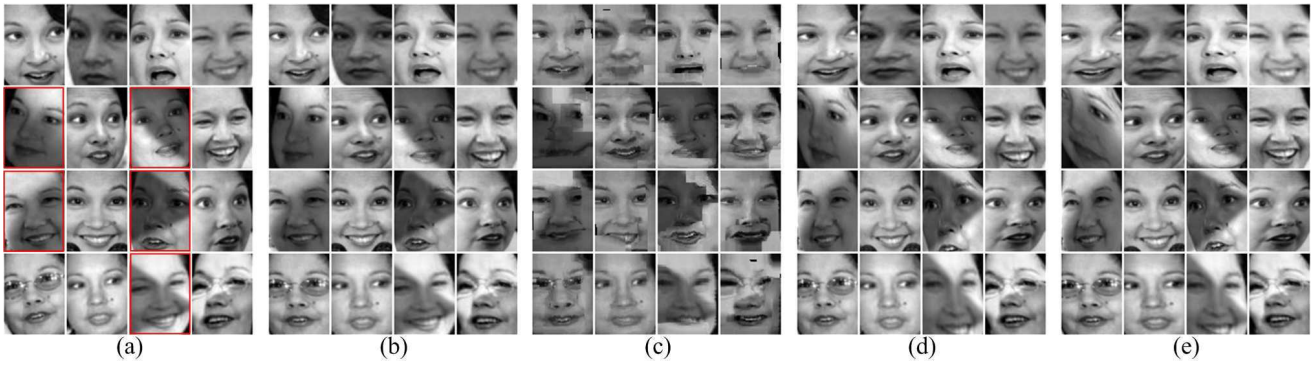

Figure 7: Alignment results of a specific subject from LFW dataset. (a) Unaligned images, the images marked with red boxes are artificially occluded with shadows. (b) Well-aligned images by our occlusion- and illumination-robust method, (c) results by SIFT, (d) by RASL and (e) by t-GRASTA.

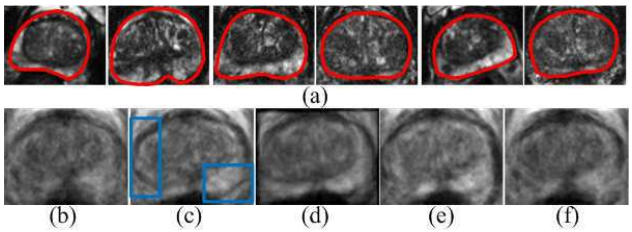

Figure 8: Alignments of MR prostate images. (a) Some unaligned prostate images from different subjects, red contours indicate prostate boundaries. The average prostate image (b) before alignment, (c) after alignment by our method, (d) by SIFT, (e) by RASL and (f) by t-GRASTA.

Alignment	SIFT	RASL	t-GRASTA	Ours
Rec. rate(%)	52.28	88.42	81.40	**91.56**

Table 1: Recognition rates on the LFW dataset for different alignment methods and SRC.

The recognition rate of SRC using our alignment method outperforms all the methods with the same "no outside data for training and testing" protocol, except for one [1] that further extracts series of discriminative features from aligned images for recognition. It is worth noting that our result is achieved by simply combining the proposed alignment and conventional SRC together, and both training and testing images taken from LFW dataset are unconstrained, thus our competitive recognition rate further demonstrates the robustness of our alignment method and its promise for providing more precise alignment for face recognitions. Specifically, the online mechanism of our robust alignment is beneficial to practical recognition applications, due to its ability to enrich the training set by incrementally aligning and adding newly collected training data, as well as its facility of robustly aligning testing data with the training set.

4. Conclusion

This work presents an online image alignment method that can be applied to large and increasing amount of images despite severe intensity distortions. This alignment problem is difficult since the images are dynamically increasing and there are great intensity variations (*e.g.*, illumination variations, partial occlusion and corruption) between images. To address this difficult problem, we have proposed an online image alignment method via subspace learning from image gradient orientations. We seek incremental alignment in the IGO domain such that the aligned IGO of the newly arrived image can be decomposed as the sum of an extremely sparse error and a linear composition of IGO-PCA basis learned from previously well-aligned ones. The image decomposition and alignment in the IGO domain benefits our method handling the large illumination variations and intensity distortions. We have validated the efficacy of the proposed method on extensive challenging datasets through image alignment and face recognition. The experimental results demonstrate that our algorithm can provide more illumination- and occlusion-robust image alignment than state-of-the-art methods do.

Acknowledgments

The work described in this paper was supported by the following grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 14202514 and CUHK 14203115).

References

[1] S. R. Arashloo and J. Kittler. Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features. *IEEE Transactions on Information Forensics and Security*, 9(12):2100–2109, 2014. 8

[2] A. Asthana, S. Zafeiriou, G. Tzimiropoulos, S. Cheng, and M. Pantic. From pixels to response maps: Discriminative image filtering for face alignment in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1312–1320, 2015. 1

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011. 3

[4] J. Eckstein and W. Yao. Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. *RUTCOR Research Reports*, 32, 2012. 3

[5] Z. Gao, L.-F. Cheong, and Y.-X. Wang. Block-sparse rpca for salient motion detection. *IEEE transactions on pattern analysis and machine intelligence*, 36(10):1975–1987, 2014. 1

[6] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6):643–660, 2001. 2

[7] J. He, D. Zhang, L. Balzano, and T. Tao. Iterative online subspace learning for robust image alignment. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2013. 2, 3, 4, 5, 7

[8] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 1

[9] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002. 4

[10] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua. Labeled faces in the wild: A survey. In *Advances in Face Detection and Facial Image Analysis*, pages 189–248. Springer, 2016. 5, 7

[11] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2006. 1

[12] Y. Li, C. Chen, and J. Yang, Fei and. Deep sparse representation for robust image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4894–4901, 2015. 1

[13] Z. Li, D. Mahapatra, J. A. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos. Image registration based on autocorrelation of local structure. *IEEE transactions on medical imaging*, 35(1):63–75, 2016. 1

[14] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):978–994, 2011. 5

[15] T.-H. Oh, H. Kim, Y.-W. Tai, J.-C. Bazin, and I. So Kweon. Partial sum minimization of singular values in rpca for low-level vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 145–152, 2013. 1

[16] N. Parikh, S. P. Boyd, et al. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014. 3

[17] X. Peng, S. Zhang, Y. Yang, and D. N. Metaxas. Piefa: Personalized incremental and ensemble face alignment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3880–3888, 2015. 2

[18] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. 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. 1, 2, 5, 6, 7

[19] G. Puglisi and S. Battiato. A robust image alignment algorithm for video stabilization purposes. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(10):1390–1400, 2011. 1

[20] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3):125–141, 2008. 4

[21] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 1003–1011. IEEE, 2015. 2

[22] W. Song, J. Zhu, Y. Li, and C. Chen. Image alignment by online robust pca via stochastic gradient descent. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(7):1241–1250, July 2016. 2, 4

[23] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013. 1, 7

[24] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006. 1

[25] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. Subspace learning from image gradient orientations. *IEEE transactions on pattern analysis and machine intelligence*, 34(12):2454–2466, 2012. 3, 5

[26] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 5

[27] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, 2012. 1, 7

[28] Y. Wang, J.-Z. Cheng, D. Ni, M. Lin, J. Qin, X. Luo, M. Xu, X. Xie, and P. A. Heng. Towards personalized statistical deformable model and hybrid point matching for robust mr-trus registration. *IEEE transactions on medical imaging*, 35(2):589–604, 2016. 7

[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009. 7

[30] Y. Wu, B. Shen, and H. Ling. Online robust image alignment via iterative convex optimization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1808–1814. IEEE, 2012. 2, 3

[31] T. Zhang, S. Liu, N. Ahuja, M.-H. Yang, and B. Ghanem. Robust visual tracking via consistent low-rank sparse learning. *International Journal of Computer Vision*, 111(2):171–190, 2015. 1