Approximated RPCA for Fast and Efficient Recovery of Corrupted and Linearly Correlated Images and Video Frames

Salehe Erfanian Ebadi, Ebroul Izquierdo
School of Electronic Engineering and Computer Science
Queen Mary University of London, London E1 4NS, United Kingdom
s.erfanianebadi@qmul.ac.uk

***Abstract*—This paper presents an approximated Robust Principal Component Analysis (ARPCA) framework for recovery of a set of linearly correlated images. Our algorithm seeks an optimal solution for decomposing a batch of realistic unaligned and corrupted images as the sum of a low-rank and a sparse corruption matrix, while simultaneously aligning the images according to the optimal image transformations. This extremely challenging optimization problem has been reduced to solving a number of convex programs, that minimize the sum of Frobenius norm and the ℓ_1-norm of the mentioned matrices, with guaranteed faster convergence than the state-of-the-art algorithms. The efficacy of the proposed method is verified with extensive experiments with real and synthetic data.**

***Keywords*—Image alignment, robust principal component analysis, low-rank, sparse, video stabilization.**

I. Introduction

In recent years, the popularity of image and video sharing websites such as Facebook, Instagram, YouTube, etc. has led to a dramatically large amount of data becoming available online. Applications such as face, digit, and object recognition is a problem domain in computer vision where low-dimensional linear models have received a great deal of attention. The available substantial data can be very challenging (if not impossible) to process with computer vision algorithms, if the difficulties such as significant illumination variation, occlusion, misalignment, deformities, and noise are not dealt with using a proper method. The most challenging task is aligning a set of images of an object to a fixed canonical template, simultaneously with removing occlusions, corruptions, and specularities to obtain an accurate representation of the object of interest based on similarity, for robust recognition or classification.

A great deal of progress has been made in batch image alignment, the most notable of which is [1], where the authors used a similar convex relaxation program in which the transformed images of an object from a set of unaligned images were decomposed as the sum of images from a low-rank approximation, and sparse large errors. Their algorithm was successful in cases of rigid and parametric classes of transformations, given the amount of misalignment and corruption was within a limited bound, and image sizes were not too big. While their method demonstrated robustness to corruption and occlusion, it uses a very expensive optimization program, based on a Lagrangian multipliers iterative linearization, whose performance is slow in applications where real-time or very fast performance is sought. Another work [2], minimizes a rank surrogate, however lacks robustness to corruption and occlusion. In addition, the canonical frame that their algorithm could handle was a small image of 49×49 pixels with only a small Euclidean transformation and limited corruption.

In this paper, a new algorithm is introduced for recovery of linearly correlated images and video frames (or signals), despite occlusions, corruptions, and large misalignment. Our method builds on recent advances in rank minimization and formulates the problem of batch image alignment as the solution of a subproblem in the sequence of convex programs. The solution of these convex programs have been shown to be efficient in our preceding work [3]. Our algorithm can handle batches of high resolution (up to HD quality) images in several minutes. We verify the efficacy and accuracy of our algorithm as well as its superiority to similar methods, with extensive experiments on unconstrained real images with wide range of corruption and misalignment. These results suggest the potential of our algorithm as a general tool for video stabilization, compression, and object tracking.

II. Approximated RPCA Framework

Suppose we are given n unaligned images or video frames $I_1, \ldots, I_n \in \mathbb{R}^{w \times h}$ of an object. We produce a matrix $A = [A_1, \ldots, A_n] \in \mathbb{R}^{m \times n}$ by concatenating all elements of I in row-order as columns of A. The matrix A should then be *low-rank* – since its columns are linearly correlated – with a low-rank component L, and the large errors can be expressed as the sum of a sparse matrix S and a Gaussian noise matrix G, while the parametric transformations τ can model the potential global misalignment.

$$A \circ \tau = L + S + G \tag{1}$$

$A_j \circ \tau_j$ denotes the j-th frame after transformation parameterized by the vector $\tau_j \in \mathbb{R}^\rho$ where ρ is the number of parameters fully describing the global motion model. Therefore $\rho = 4$ corresponds to similarity, $\rho = 6$ to affine, and $\rho = 8$ to projective transformation. It was shown that for the problem of recovering low-rank matrices from sparse errors, as long as the rank of the matrix A to be recovered is not too high and the number of the errors is not too large, minimizing the natural convex surrogate for $rank(A) + \lambda\|S\|_0$ (with λ soft-thresholding parameter) can *exactly* recover A [4]. In this paper, we use a different convex relaxation that replaces $rank(\cdot)$ with the *Frobenius norm*: $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$, and the ℓ_0-norm $\|S\|_0$ with the ℓ_1-norm: $\|S\|_1 = \sum_{i,j} |S_{ij}|$ in an approximated noisy case. Applying this relaxation to (1) yields a new optimization problem, such that $A \circ \tau \approx L + S$:

$$\underset{\substack{L,S,\tau \\ rank(L) \leq k}}{\arg\min} \ \|A \circ \tau - L - S\|_F + \lambda\|S\|_1 \tag{2}$$

The authors in [5] showed that for convex, Lambertian objects, images taken under distant illumination lie near an approximately nine-dimensional linear subspace known as the *harmonic plane*. However, with face images which are neither perfectly convex nor Lambertian, this low-rank model is violated, due to cast shadows, specularities, occlusions, and misalignment. These errors are large in magnitude, but sparse in the spatial domain. Given a sufficient number $n > rank(A)$ of those images, the extremely efficient and computationally inexpensive approximated Robust Principal Component Analysis in (2) will be able to remove those errors, as well as align all those images in the same canonical template. To solve this problem we use an alternating strategy minimizing the function for three parameters L, S, and τ one at a time until convergence; for a fixed λ the iterative process below will have a monotonically decreasing value, converging to a local minimum:

$$\tau^t = \arg\min_{\tau} \|A \circ \tau - L^{t-1} - S^{t-1}\|_F^2 \tag{3}$$

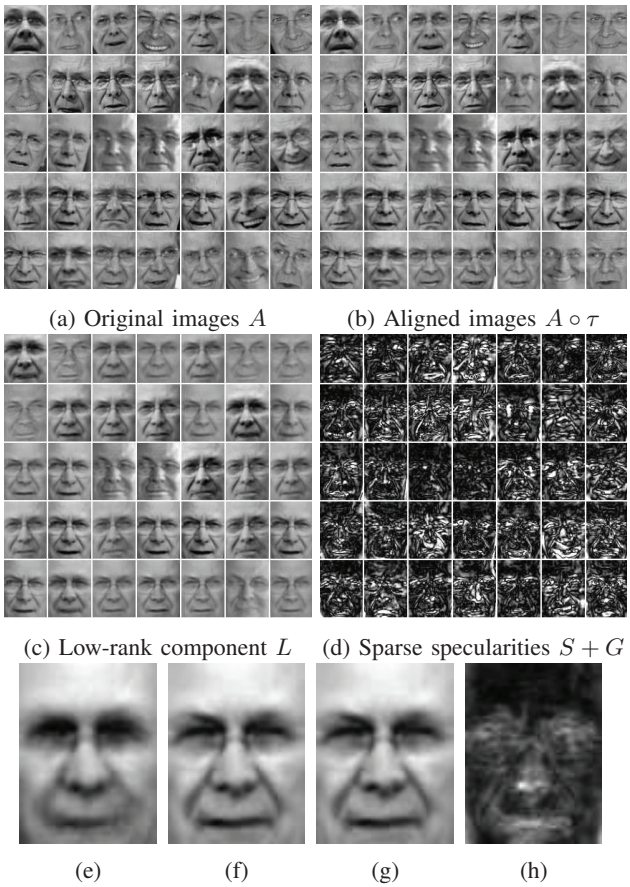
$$L^t = \arg\min_{rank(L)\leq k} \|A \circ \tau^t - L - S^{t-1}\|_F^2 \tag{4}$$

$$S^t = \arg\min_{S} \|A \circ \tau^t - L^t - S\|_F^2 + \lambda\|S\|_1 \tag{5}$$

The main remaining difficulty in solving (2) is the non-linearity of the constraint $A \circ \tau \approx L + S$, which arises as a result of the dependence of $A \circ \tau$ on the transformations τ. We use the linearization method described in [3], where an incremental refinement is used. The i-th geometric transformation is comprised of a parameter vector τ_i, $i = 1, \ldots, n$ where different spatial transformations can be considered. We use the $2D$ parametric transforms to model the translation, rotation, and planar deformation in the low-rank subspace. We obtain an initial approximation for the parameters τ_i using a feature matching, indirect method with SIFT features [6] where the images are aligned to the middle image. This method is more robust and much faster compared to direct methods used in [1] with larger image sizes and more extreme parametric transformations and large camera parallax, displacement, and motion blur. Finally, in (3) we use the multi-resolution incremental refinement described in [7], to estimate these motion parameters. To calculate the rank-k matrix that is the nearest estimate of the matrix $A \circ \tau^t - S^{t-1}$ in (4), SVD gives a closed-form solution as: $L^t = \sum_{i=1}^{k} \sigma_i U_i V_i^T$, with the coefficients σ_i the singular values, and the vectors U_i and V_i the singular vectors of the matrix $A \circ \tau^t - S^{t-1}$. Finally in (5) the matrix S^t is updated using the parameter λ acting as a regulating parameter, where the elements of the matrix $A \circ \tau^t \leq \lambda$ are considered zero.

III. Experiments

In this section, we demonstrate the efficacy of our method in a variety of image recovery tasks. We verify the correctness of our method with controlled and uncontrolled examples, and show that it outperforms state-of-the-art methods in recovery of corrupted data while simultaneously compensating for any misalignment. Our realistic examples are taken from the challenging Labeled Faces in the Wild (LFW) database [8]. Experiments on video data and handwritten digits further indicate the generality of our method for various applications. Moreover, our algorithm can handle more complicated deformations and transformations such as planar homographies as shown in one of the tests, which indicates wide range of applications in video stabilization and compression.



(a) Original images A (b) Aligned images $A \circ \tau$

(c) Low-rank component L (d) Sparse specularities $S + G$

(e) (f) (g) (h)

Figure 2: Robust alignment by sparse and low-rank decomposition in LFW dataset [8]. Contrast has been normalized in (d) for better visualization. Figures (e), (f), (g), and (h) correspond to the average of (a), (b), (c), and (d) respectively.

A. Speed and scalability of our method

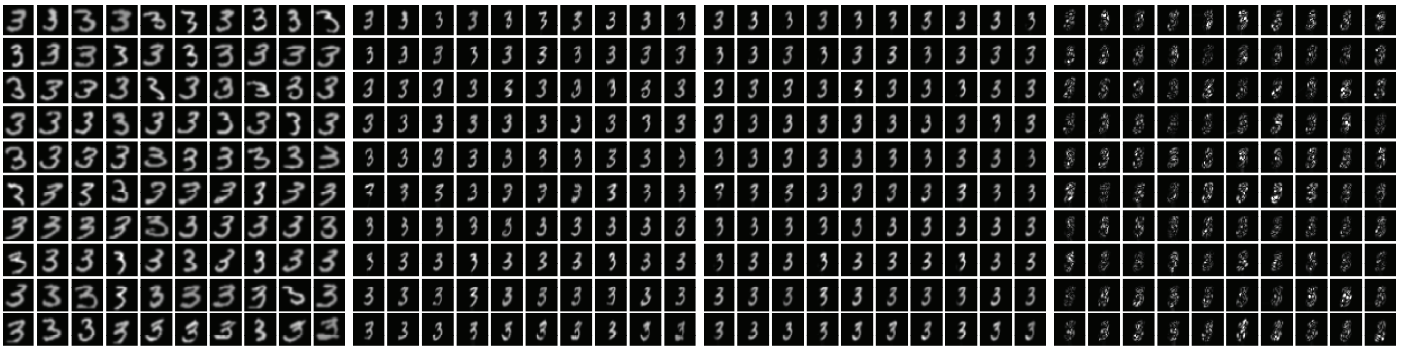
For this example, on a 3.40GHz (single core) Intel Core i7-4770 machine with 32GB of RAM our Matlab implementation requires 11.07 seconds to recover and align 100 perturbed and corrupted synthetic images of size 49×49, whereas [1] requires 41.44 seconds. Moreover, our algorithm is able to handle large image sizes (up to HD quality), which demonstrates impressive computational efficiency as a direct result of using our approximated RPCA optimization framework.

B. Removing shadows and specularities from face images

We test our algorithm using a set of controlled images. Figure 1 shows 100 images of a dummy head that are perturbed and occluded randomly. The images are all 49×49 pixels (our algorithm can handle much larger image sizes, however for comparison with similar methods the same image data have been used). To each image a random Euclidean transform is applied with angle of rotation uniformly distributed in $[-10\,^\circ, 10\,^\circ]$ and x- and y-translations are uniformly distributed in $[-3, 3]$ pixels, while 6% of the pixels are corrupted.

(a) Original corrupted images A (b) Aligned images $A \circ \tau$ (c) Low-rank component L (d) Sparse specularities $S + G$

Figure 1: Robust alignment by sparse and low-rank decomposition in Synthetic face data.



(a) Original handwritten digits A (b) Aligned digits $A \circ \tau$ (c) Low-rank component L (d) Sparse errors $S + G$

Figure 3: Robust recovery and alignment by sparse and low-rank decomposition in handwritten digits.

Notice that our method correctly removes the occlusions (Figure 1-(d)), to produce a rank 3 matrix of well-aligned images (Figure 1-(c)). RASL [1] can produce the same results but with the minimized rank 48. The rank 3 matrix best describes the general appearance of the face image in this case, while preserving the prominent features for recognition purposes.

Next, we validate our approach using more challenging images taken from Labeled Faces in the Wild (LFW) [8] dataset of public figures. These images exhibit significant variations in pose and facial expression, illumination, and occlusion; moreover the ground truth (i.e. undistorted, not rotated, not shifted) image is not known. The images are aligned to a 80×60 canonical frame, and Affine transformations are used to cope with large variability in poses. Figure 2 shows one example from this dataset. Notice the average face after alignment is significantly clearer in Figure 2-(f) indicating improved alignment achieved by our method. This example demonstrates our method's ability in correcting errors in real images, which could be used to improve the performance of current face recognition systems.

C. Recovery of corrupted and misaligned handwritten digits

Our method can be applied to aligning any general set of images with strong linear correlation. In this test, we used 100 handwritten digits "3" from the MNIST database in Figure 3. Our algorithm can obtain comparably good performance on this example despite the fact that it does not explicitly target binary image alignment.

D. Recovery of deformed and corrupted planar surfaces

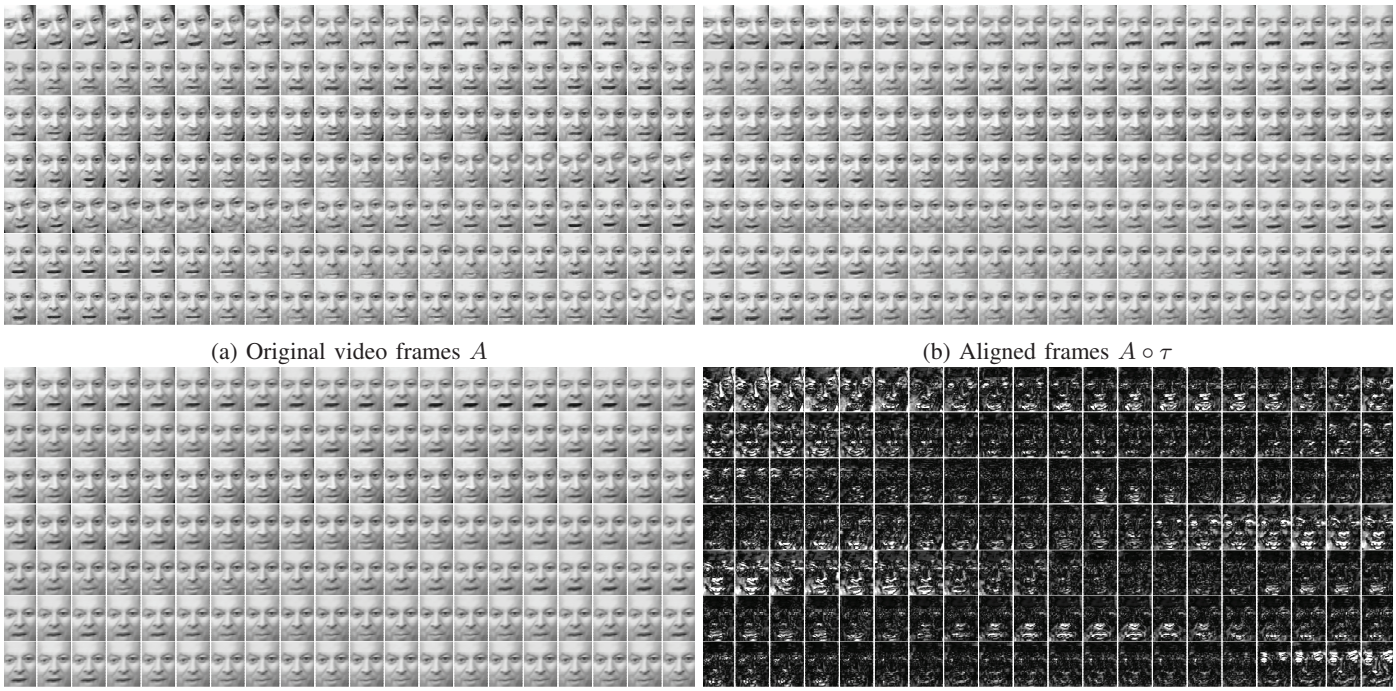
In this example, our algorithm is applied to images that differ by planar homographies, to demonstrate how it can be used with more complicated deformation models. Figure 4 shows 8 images of a building, taken from various viewpoints by a perspective camera. As seen here, the algorithm correctly aligns the windows and removes branches occluding them. This hints a useful application for our method in image matching, mosaicing, and inpainting.

E. Video stabilization for recovery of object of interest

Video frames taken from the same scene are usually linearly correlated. In this test, we demonstrate the ability of our method in aligning frames taken from a video. Figure 5 shows frames from a 140 frame video of Al Gore talking, obtained by applying a face detector to each frame individually. Due to imprecision in face detector there is high jitter from frame to frame. Next, we use affine transformations to obtain a well-aligned set of frames, and then we demonstrate a low-rank approximation of the frames as well as the removed shadows, occlusions, and errors from the images. Notice that the errors shown in Figure 5-(d) compensate for local motion such as mouth movements, and eye blinking which are not considered in the global motion model. For this video of 80×60 pixels with 140 frames our method needs 9.79 seconds while [1] takes 57.52 seconds to produce visually similar results. These results suggest the potential of our algorithm as a general tool for video stabilization, compression, and object tracking.

(a) Original homography images A (b) Aligned images $A \circ \tau$ (c) Reconstructed images L (d) Removed occlusions $S + G$

Figure 4: Alignment and recovery of planar homographies.



(a) Original video frames A (b) Aligned frames $A \circ \tau$

(c) Aligned and recovered images L (d) Removed errors $S + G$

Figure 5: Video stabilization for recovery of object of interest.

IV. Conclusion

In this paper we demonstrated the surprising effectiveness and efficacy of our approximated RPCA method for batch image recovery from corruptions and misalignment, and suggested applications such as batch image alignment, recovery of face images form corrupted data for face recognition, video stabilization, image mosaicing, inpainting etc. Our proposed formulation directly impacts the speed of convergence of the algorithm, making it suited for real-time applications. One of the most important questions for future work is how to extend our framework to more general classes of transformations such as non-rigid and non-parametric that are exhibited in general video data, while providing the same theoretical guarantees for the amount of misalignment and corruption it can handle.

V. Acknowledgement

This work is supported in part by the LASIE project (http://www.lasie-project.eu/) with funding from the European Unions Seventh Framework Programme for research, technological development.

References

[1] 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*, vol. 34, no. 11, pp. 2233–2246, 2012.

[2] A. Vedaldi, G. Guidi, and S. Soatto, "Joint data alignment up to (lossy) transformations," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[3] S. Erfanian Ebadi, V. Guerra Ones, and E. Izquierdo, "Efficient background subtraction with low-rank and sparse matrix decomposition," in *Image Processing (ICIP), 2015 IEEE International Conference on*, September 2015.

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

[5] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, pp. 218–233, 2003.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010.

[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.